

Come veto or high water

The obduracy of the White House will slow the progress of stem-cell research in the United States — just as Europe agrees to move forward with it.

President Bush's decision to veto legislation that would allow the federal government to support human embryonic stem-cell research is a monumental error. The decision contrasts with a broadly satisfactory compromise reached by the European Union (EU), which opens the way for more funding of such research across Europe (see page 335).

Particularly in the light of the EU decision, the first veto of the Bush presidency will enrage US biologists. But it should not be allowed to obscure the fact that the public debate over the ethics of this research is being won. Over the past two years, the campaign to win public support for stem-cell research in the United States has been successful. Some two-thirds of the American public now support the work. The bill that Bush vetoed was passed by healthy majorities in both houses of Congress, as more and more conservative law-makers have come round to the idea that the work should proceed under the auspices of the National Institutes of Health.

In Brussels, a compromise was reached on 24 July whereby the EU will stop paying for the derivation of stem cells from embryos, but will support research using these cells. With customary opaqueness, the 25 research ministers did not cast a vote at the meeting, but instead hatched a practical and welcome deal that will allow this important work to proceed.

In the United States, however, there is no prospect of even the new Congress elected this November mustering the two-thirds majority needed to overturn the veto. It will be 2009 at the earliest before a president willing to support this work enters the White House — eight long years after the current Bush policy first came into effect. The scientific opportunities squandered in that time are irretrievable; the years of human life and health lost, unknowable.

The end of the line

The bill passed by Congress would have overturned the policy, announced by the president on 9 August 2001, that allows federally funded research to be done only on embryonic stem-cell lines created before that date. Bush said at the time that more than 60 such lines were available. But many of those lines proved impossible to grow, and only 21 lines are ready today for federally funded research.

Some of these are of limited use in basic research because they have accumulated genetic mutations over time, and they are all less than ideal for development into clinical treatments because mouse cells and serum were used to grow them, raising the remote but real possibility that they may carry disease. Access to additional lines containing abnormalities associated with diseases, such as Huntington's, would allow the study of how the disease occurs and, perhaps, lead to therapies not based on cell or tissue transplantation.

The vetoed bill would have allowed federal money to support research on new cell lines derived from thousands of embryos left over at *in vitro* fertilization clinics and currently slated for destruction.

It would only permit these embryos to be used in research after their parents had independently decided that they did not want to implant them. According to a RAND Corporation study of some 400,000 embryos in storage, parents have consented to donate about 11,000 of these for research, from which perhaps 275 new lines could be derived (D. I. Hoffman *et al. Fertil. Steril.* **79**, 1063–1069; 2003).

But it is not just an order-of-magnitude increase in the number of lines — and the power of the federal purse to finance their exploitation — that the law would have brought to bear. It would have brought the daylight, research coordination and ethical oversight that accompany a much wider federal presence in the field.

Some state governments have stepped into the funding breach: the day after the Bush veto, California governor Arnold Schwarzenegger announced that he would provide \$150 million to jump-start California's \$3-billion stem-cell initiative. But a patchwork of state, university and foundation-financed activities is no substitute for a properly coordinated effort led by the National Institutes of Health. Additionally, precious time, energy and money is being wasted building separate facilities and purchasing duplicative equipment for work on stem-cell lines that have not been approved for federal funding.

It can be argued in good faith that not a single embryo should be destroyed in the name of medical progress. However, Bush has done nothing to prevent the wholesale destruction of embryos in *in vitro* fertilization clinics. And the arguments he used in a speech announcing the veto were disingenuous. If the bill became law, he said, "American taxpayers would, for the first time in our history, be compelled to fund the deliberate destruction of human embryos. And I'm not going to allow it."

This is patently untrue. The vetoed bill would not have allowed federal funding for the derivation of human embryonic stem cells from leftover embryos at fertility clinics, and the attendant destruction of the embryos. It simply allowed taxpayer-funded research on the stem cells thus derived. The derivation itself would still need to be financed with non-federal funds.

This may seem like a subtle point, but it exactly this kind of compromise that has forged America's uneasy but workable abortion policy: abortions are legal, but in no case are taxpayers required to see their dollars fund them. It may be tenable to argue that dissenters should not be asked to finance a practice that they find morally unacceptable. What is not acceptable is for the president to use false pretences to stand in the way of a compromise that the Congress has sensibly endorsed.

"The vetoed bill would have allowed federal money to support research on new cell lines derived from thousands of embryos currently slated for destruction."



Still not alert

Tsunami preparations in the Indian Ocean remain inadequate.

At the end of June, the United Nations Educational, Scientific and Cultural Organization (UNESCO) announced that the Indian Ocean Tsunami Warning System was “up and running”. Yet on 17 July, a tsunami in Indonesia killed more than 550 people. These events demonstrate a continuing disconnect between the rhetoric of international organizations and the reality on the ground.

In contrast to the Indian Ocean tsunami that killed almost 300,000 people in December 2004, Indonesian authorities this time received a warning from the Pacific Tsunami Warning Center. But the government did nothing with the information, fearing, according to Kusmayanto Kadiman, Indonesia’s science minister, that raising a false alarm would cause unnecessary panic.

Indonesia’s response to the latest disaster reflects the fact that, in the rush for action that followed the 2004 tsunami, too much emphasis has been put on the expansion of the high-tech early warning system, and not enough on improving local preparedness.

This tsunami hit the Indonesian island of Java in the afternoon, when beaches were crowded with tourists — about 40 minutes after the warning was issued. That left plenty of opportunity for evacuation, if there had been sirens, for example, or if muezzins in the local mosques were ready to warn people, or if the population knew better how to recognize the ground or sea movements that presage a tsunami.

The problem is by no means confined to Indonesia. Earthquakes as far apart as Tonga and Greece earlier this year might each have generated tsunamis, but in neither case were warnings issued to the public. In Tonga the authorities blamed an ill-functioning fax machine.

The managing board of the Indian Ocean warning system, which meets next week in Bali, must continue to implement its still-incomplete system — but must also concentrate on ensuring that countries in the region get the right advice on how to make use of such warnings.

When the Java tsunami struck, it took the Pacific Ocean warning system only a few minutes to respond with a warning. That time could be reduced marginally if there were more seismic stations, but what really matters now is improving national emergency management in countries such as Indonesia. UNESCO must press governments to implement better disaster planning, and, wherever possible, it should get more actively involved in tackling inadequacies itself.

If more local scientists were trained in tsunami modelling and forecasting they could advise local communities on simple actions to protect residents from tsunami-generating earthquakes. Some 60 scientists from 18 Indian Ocean nations have, for example, attended courses in Hawaii on how to produce inundation maps for warning guidance. It is through initiatives like that — ones that build local expertise from the ground up — that the death toll from future tsunamis may be minimized. ■

“Too much emphasis has been put on the expansion of the high-tech early warning system, and not enough on improving local preparedness.”

Let’s replicate

Post-publication follow-up evolves.

The replication of research results is a linchpin of the scientific process. But it isn’t always done the way a layperson might expect — by replicating a previous experiment, step-by-step, in a different lab. Instead, validation of research results can take different forms in different disciplines. And as we report on page 344, changes in scientific publishing could herald a new era of open, interactive communication in which research findings are tested.

The practice of replication is already more diverse, and less clean-cut, than is commonly realized. Funding agencies won’t normally pay for the direct repetition of published work. In some disciplines, fresh work that builds on a published result will involve fully replicating the work that led to that result. In others, including many branches of biology, scientists see the process as validation, rather than replication. The expectation isn’t necessarily that another scientist will reproduce exactly what they did, but that the data and methodology are strong enough to withstand inferences and have new experiments built upon them.

Whereas outside observers may think that all science can, and should, be reproduced, most scientists aim to confirm and extend published conclusions. That is a significantly different objective, and

may be accomplished most compellingly using a different experimental approach, in pursuit of the answer to a related but separate question. This allows investigators to get around the practical difficulties of replication, such as the fact that materials or animals are never truly identical. Because of these difficulties, failed attempts to perform an exact replication of the data from a multivariable experimental system do not automatically invalidate the conclusions of the original study.

However, the interrogation of previously published results remains an essential part of science. Too often in the past, failure to achieve successful replication has left unsatisfactory research in a kind of helpless limbo. Such non-replication may be the subject of late-night arguments at scientific meetings, but is unlikely ever to be published, or even submitted for publication. Indeed, in many disciplines, it is difficult to raise concerns about an initial result in a public forum without this being construed as a full-frontal assault on the original authors’ integrity.

But publishing is changing, and the idea of a scientific paper as a self-contained and unalterable body of work is gradually becoming obsolete. In some disciplines (physicists have led the way, through arXiv), scientific results are already subject to what amounts to an open-ended peer-review process, in which colleagues can continue to criticize and discuss the contents of a paper, for as long as it takes. That, rather than the current, sometimes awkward silence, is the way ahead. ■

RESEARCH HIGHLIGHTS

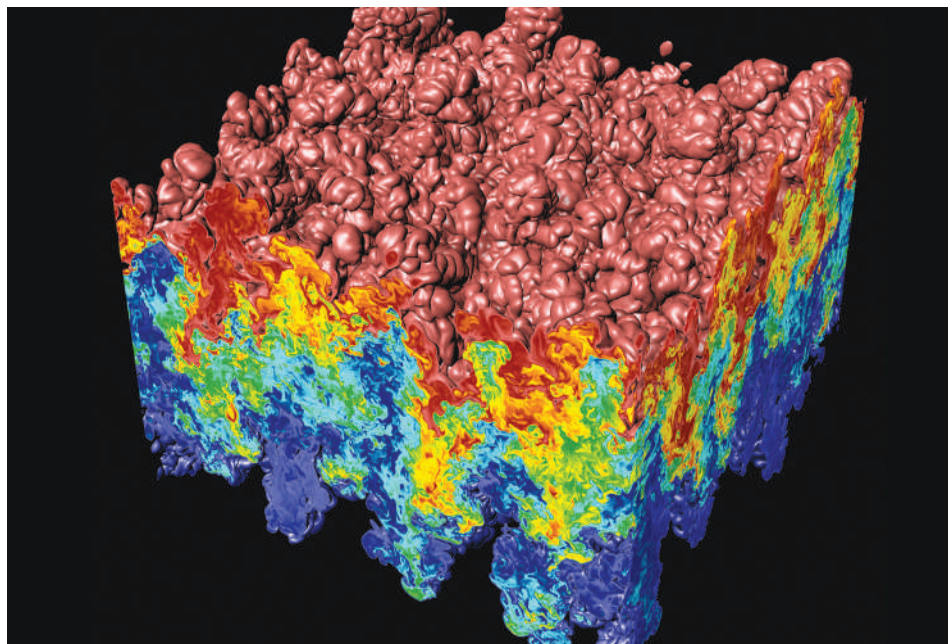
Top heavy

Nature Phys. doi:10.1038/nphys361 (2006)

The swirling motion of turbulence has been simulated in unprecedented detail using the world's fastest supercomputer.

William Cabot and Andrew Cook of the Lawrence Livermore National Laboratory, California, used IBM's BlueGene/L to model what happens when a layer of heavy fluid (shown in red) sits above a lighter fluid (blue). This situation is described as Rayleigh–Taylor unstable: buoyancy effects mean that any bumps in the lower layer grow, setting up eddies that multiply until the entire fluid is seething (pictured).

Cabot and Cook report the surprise finding that the rate at which the eddies grow and the fluids mix depends on a property of the fluids known as the Reynolds number. This has implications for models of supernovae, which are turbulent stellar explosions.



CHEMISTRY

Solvent speed limits

Angew. Chem. Int. Edn **45**, 4824–4825 (2006)

While singing the praises of ionic liquids as 'green' solvents for industrial chemistry, advocates should not overlook the virtues of the original green solvent — water.

Ionic liquids are more environmentally friendly than organic solvents because they are not volatile. But Shraeddha Tiwari and Anil Kumar of the National Chemical Laboratory in Pune, India, have shown that an important class of ring-forming reactions in synthetic organic chemistry, known as Diels–Alder reactions, proceed faster in water than in ionic liquids. It was known already that Diels–Alder reactions can proceed faster in water than in organic solvents, but the new results clinch the case for water being the solvent of choice for this type of chemistry.

NEUROBIOLOGY

Monkey talk

Nature Neurosci. doi:10.1038/nm1741 (2006)

Brain-imaging studies performed on rhesus macaques are helping to trace the evolutionary origins of human speech.

Allen Braun of the National Institutes of Health in Bethesda, Maryland, and his colleagues monitored blood flow in the brains of three rhesus macaques as they listened to recordings of the coos and screams made by monkeys of their species. The flow of blood increased in regions of the

brain similar to the principal human language centres. The results complement those of an earlier study (A. Poremba *et al.* *Nature* **427**, 448–451; 2004) that had focused on the temporal lobes of monkeys' brains.

These findings suggest that the evolution of language in humans may have involved the recruitment of prelinguistic centres in the common ancestor of macaques and humans more than 25 million years ago.

PHYSICS

Beating time

Phys. Rev. Lett. **97**, 020801 (2006)

Scientists have used a single mercury atom to build the world's most precise clock. Its time-keeping surpasses that of the caesium atomic clocks currently used to define the second.

The beat of the clock, built by Jim Bergquist and his colleagues at the National Institute of Standards and Technology in Boulder, Colorado, is defined by the oscillations of laser light that excite the mercury atom. Using an ultraviolet laser makes the clock about five times more precise than its caesium counterparts, which use a lower-frequency microwave laser. It is also more reliable: the clock is expected to drift by only 1 second in 400 million years compared with 1 in 60 million in caesium clocks.

According to the team, the mercury clock could be used to measure fundamental physics constants and, ultimately, to redefine the second.

PHARMACEUTICALS

Change of heart

Nature Med. doi:10.1038/nm1446 (2006)

A cancer drug hailed for its ability to rescue those dying from leukaemia could cause heart failure, a study suggests.

Imatinib mesylate (Gleevec) was carefully designed to target the tyrosine-kinase enzyme mutated in chronic myelogenous leukaemia. Thomas Force at Jefferson Medical College in Philadelphia, Pennsylvania, and his colleagues document ten patients with otherwise healthy hearts who developed heart failure after taking the drug. Gleevec, produced by Novartis, seems to stress a membrane system known as the endoplasmic reticulum in heart muscle cells, and can eventually trigger cell death.



Patients taking Gleevec should be monitored for heart problems, the authors say, as should those in clinical trials of the many tyrosine-kinase inhibitors under development.

BIOPHYSICS

Gibbons bounce about

J. Exp. Biol. **209**, 2829–2838 (2006)

Gibbons, known for their graceful skill at swinging from tree to tree, also occasionally travel by foot. Evie Vereecke and her two co-authors at the University of Antwerp, Belgium, have investigated this ground-based locomotion by filming white-handed gibbons (*Hylobates lar*, pictured right) and recording the pressure exerted by the gibbons' feet on a special mat at various speeds. Unlike humans, who walk and then break into a run, the gibbons' gait transition is not clear cut.

Walking humans swing their legs like pendulums, whereas the legs of people running act like springs, using their Achilles tendons to store energy. The researchers found that gibbons have a bouncing gait that works at all speeds, probably storing energy in their quadriceps muscles. Running gibbons reach speeds of 3.5 metres per second, but rarely have both feet off the ground at the same time.

CHEMISTRY

Gold winner

Science **313**, 332–334 (2006)

Avelino Corma and Pedro Serna of the Technical University of Valencia, Spain, have discovered a clean and selective way to make aromatic amines, compounds that are crucial to the production of herbicides, pharmaceuticals, dyes and pigments.

The researchers showed that gold nanoparticles, fixed to a titanium dioxide or iron oxide substrate, could catalyse the exchange of oxygen atoms for hydrogen in a nitro group (NO_2), forming an amine (NH_2). This chemical reduction, using hydrogen gas, proceeded without affecting other groups in the same molecule. Previous approaches required difficult-to-remove and toxic additives, such as the soluble heavy metal, vanadium. This new reaction also avoids the build-up of explosive intermediates that troubles some other schemes.

CELL BIOLOGY

Gene control on an RNA diet

Nature Cell Biol. doi:10.1038/ncb1439 (2006)

Clues to how cells gather gene-censoring molecules from their environment may lead to novel therapeutic strategies, suggest Raul



R. MAGNUSSEN/ALAMY

Andino of the University of California, San Francisco, and his colleagues.

The researchers studied the mechanism by which cells from the fruitfly *Drosophila* take up double-stranded RNA. Once in the cell, this RNA is chopped into pieces to silence the expression of specific genes, by a mechanism known as RNA interference.

They identified 23 genes involved in the RNA uptake, including some linked to endocytosis — the process by which external molecules are enclosed in vessels called vesicles. The team speculates that a family of 'pattern recognition receptors' detect the RNA, triggering its consumption.

MATERIALS SCIENCE

Property conversion

J. Am. Chem. Soc. doi:10.1021/ja0627996 (2006)

Even inert metals tend to have reactive surfaces, so nanocrystals with a large surface-area per unit volume can be persuaded to react in ways that bulkier bits of material will not.

Robert Cable and Raymond Schaak of Texas A&M University, College Station, have applied this principle to a class of metal alloys known as intermetallics, which have chemical properties distinct from those of their constituent metals. Heating intermetallic nanocrystals with solutions of metal salts yielded compounds that can be difficult to synthesize by other routes, the researchers report.

They showed, for example, that it was possible to fine-tune the composition of a nanocrystal containing platinum and tin through a reversible conversion between different alloys. Such materials are being tested as catalysts for fuel-cell reactions.

JOURNAL CLUB

Thomas Koop
Bielefeld University, Germany

A physical chemist describes how ice cream might be kept smooth for longer.

I am addicted to ice cream, so it is always a let-down to come across an old batch that has become gritty. The unpleasant texture develops as large ice crystals grow at the cost of smaller ones in a recrystallization process known as Ostwald ripening. Is there a way to prevent this happening?

Ice recrystallization is also a threat to organisms that live at subzero temperatures, because intracellular ice growth can damage their tissues. In response, plants, fish and insects synthesize antifreeze proteins. Such proteins inhibit recrystallization by adsorbing to certain faces of the ice crystals.

A recent article (Li *et al.* *J. Chem. Phys.* **124**, 204702; 2006) presents a new and promising model for understanding how the microscopic interactions between ice and protein translate to observable effects — a subject much discussed.

The model calculates how much of the surface of the ice is covered by protein by describing a dynamic equilibrium between unbound and adsorbed protein. This depends on the concentration of the protein in solution, its size, the number of potential binding sites the protein has and its binding strength.

The model predicts the number of degrees below the equilibrium melting temperature at which the protein will inhibit ice growth. Its predictions are in good agreement with experimental data for several proteins.

A better understanding of the ice-antifreeze protein interaction should speed up development of mimetics for use in the cold storage of biological tissues or frozen food. Until then, I will use Ostwald ripening as a scientific argument for finishing off each batch of freshly prepared homemade ice cream in one go. I don't need to argue with my kids about this, as they follow the strategy intuitively.

NEWS

Nigeria ready for huge science spend

The politics of Nigeria's government, known for its corruption and squandering of oil money, could diplomatically be described as hard to predict. But if plans due to be discussed by parliament later this year come to fruition, the nation's research base will undergo a dramatic transformation.

With its coffers boosted by oil exports, Nigeria is considering creating a US\$5-billion endowment fund for science and technology — an investment that would generate a research and development budget on a par with many developed nations, and bigger than those of any of its African rivals.

President Olusegun Obasanjo only gave his backing to the proposal in May, but a final decision could be taken this year, as the president is thought to be keen to push it through before leaving office next spring. "He is determined for this to be one of his legacies," says Folarin Osotimehin, a science policy adviser at the United Nations Educational, Scientific and Cultural Organization (UNESCO) in Paris.

Osotimehin is one of several experts from UNESCO and other organizations that have been working with Nigeria on the plan. The result is a proposal, expected to go before parliament this autumn, to establish a National Science Foundation based on the US organization of the same name.

The foundation would distribute the return on the endowment through peer review of competitive grant schemes, an unusual move for an African nation. Only South Africa uses competition to distribute a significant fraction of its science budget, says Mohamed Hassan, executive director of the Third World Academy of Science in Trieste, Italy, and a member of the board advising Obasanjo.

The foundation would also allow Nigeria to break with African tradition by devoting a large sum of public money to research. South Africa, currently the continent's leading science spender, has an annual budget of around \$200 million. A 10% return on the Nigerian endowment would give the country \$500 million to spend on science every year.

Whether the ambitious target of \$5 billion can actually be reached remains an open question. Foreign experts working on the plan say they are advising Obasanjo to meet the total from government funds. Thanks to Nigeria's extensive oil deposits and current high oil prices, that might be possible. But Osita Ogbu, chief economic adviser to Obasanjo, says that Nigeria will contribute the bulk of the money and approach foreign governments

and donor organizations for the rest.

That could prove a challenge, as Nigeria has long been associated with the abuse of state and donor funds. In the Corruption Perception Index, a league table compiled by the Berlin-based lobby group Transparency International, Nigeria is joint sixth from bottom of 159 nations. Oil revenue has been a particular source of controversy, with senior Nigerian government officials repeatedly being accused of siphoning off revenues.

But observers say that Nigeria is changing its ways. In 2004, for example, Obasanjo opened up the country's oil accounts to external scrutiny. The books revealed that money had been stolen in the

past, but by pledging to keep the accounts open, the government has shown its commitment to preventing future thefts, says Casey Kelso, regional director for Africa at Transparency International. He adds that more than 50 cases of serious financial crime are being processed. "There have been some hopeful developments," he concludes. "Corruption is being tackled."

For foreign donors to be convinced about the endowment and its proceeds, Nigeria will have to pledge to publish clear and comprehensive fund accounts, and to ensure that the peer review is handled by respected scientists. Kelso says that donors might also want to see restrictions put on the amount that can be taken out of the endowment every year and a cap on individual grants.

Foreign advisers say UNESCO is confident enough about the project to have offered a \$200-million loan. The money generated from the endowment is likely to be channelled into areas that Nigeria has already identified as priorities: income-generating fields such as agricultural biotechnology, information technology and satellite launch systems. "We have to see a product at the end of the project," says Ogbu. "This will be research geared towards supporting the economy."

Advisers on the project say they will finalize the details over the summer and present the result to parliament in September. Ogbu knows that these final months of Obasanjo's presidency could be the only chance to push through the plan: "It has to be set up before he leaves. Otherwise we could have a president without enthusiasm for science."

Jim Giles



Growth area: Nigeria is expected to focus funding on areas such as agricultural biotechnology.

J. SILBERBERG/PANOS

SPECIAL REPORT

A long week in stem-cell politics...

Just days after US President George W. Bush vetoed a bill that would have significantly broadened federal funding for human embryonic stem-cell research, European Union ministers have given the green light to funding guidelines similar to the US proposal.

In a cliffhanger decision on 24 July, the EU council agreed to the inclusion of ethically approved human embryonic stem-cell research in its next round of research funding. The €50-billion (US\$63-billion) seventh Framework Programme for research (FP7) is due to start in January 2007, and runs until 2013.

But the council will not directly finance the destruction of human embryos. So researchers cannot use FP7 funding to derive their own cell lines from embryos left over from *in vitro* fertilization procedures, as can be done with national funding in the United Kingdom and Sweden. They must instead buy them from other sources.

The UK Royal Society expressed “disappointment” at the restriction. But others are happier. “I think this is an OK decision given this is Europe and we have to find compromise,” says Elena Cattaneo, a stem-cell researcher at the University of Milan in Italy.

Approval of most aspects of the research programme require a qualified majority vote, and there was much horse-trading as countries with restrictive stem-cell laws, particularly Germany, tried to set up a ‘blocking minority’ of countries. These apparently failed, but only at the last minute.

Austin Smith, incoming chair of the Institute for Stem Cell Biology in Cambridge, UK, says the issue is seen as broader than just stem cells. “The whole of science is under attack with the sorts of statements George Bush has been making, and with Germany trying to impose its own moral and cultural view over the entirety of European research,” he says. “These fundamentalist positions refuse to see the benefits of science — it’s like going back to the days of Galileo and the Church. There has been a huge sigh of relief in the research community in Europe this week.”

Across the Atlantic, researchers had no such reprieve, as Bush exercised the first veto of his almost six-year presidency on 19 July. “This bill would support the taking of innocent human life in the hope of finding medical benefits for others. It crosses a moral boundary that our decent society needs to respect, so I vetoed it,” Bush declared at a White House gathering that included babies ‘adopted’ as frozen embryos left over at fertility clinics. The same day, the US House of Representatives attempted to



R. EDMONDS/AP

President Bush believes that funding broader embryonic stem-cell research would cross “a moral boundary”.

override the veto, but failed to muster the two-thirds majority needed.

The veto prompted Republican Governor Arnold Schwarzenegger of California to announce a \$150-million loan the next day to jump-start that state’s \$3-billion stem-cell research programme, which has been stalled by legal wrangling. “We cannot fall behind the nations that make this life-saving science a priority,” says Schwarzenegger.

The vetoed bill, which passed the US Senate on a vote of 63 to 37 the day before Bush’s action, would have allowed federal funding for research on stem cells derived from embryos left over at fertility clinics and already slated for destruction. As in the European agreement, it would not have allowed funds to be used to derive cell lines from such embryos. Current law forbids US funding of all embryonic stem-cell research, except on 21 lines derived before 9 August 2001.

Supporters have vowed to resurrect the legislation. “Whether it’s this year, or with a new Senate and a new House and the next president, this will become the law of the United States,” Senator Charles Schumer (Democrat, New York) told a Capitol Hill rally.

Observers and research advocates say Bush’s action could cost Republicans politically in a country where polls show that at least 60% of the public support the research. “It’s a very risky strategy for Bush,” says leading US pollster John Zogby. “It makes him and

the Republicans appear to be anti-science.”

That the legislation reached the president’s desk at all in a Congress with both houses controlled by conservative Republicans speaks to the broad reach of diseases, from juvenile diabetes to Parkinson’s, for which the research holds promise. It is also testimony to the strength of the Coalition for the Advancement of Medical Research (CAMR), an umbrella group of disease and research advocacy groups that lobbied relentlessly for the bill.

“There’s never been a group strong enough to push back against the pro-life lobby in Congress. And this group did that,” notes Arthur Caplan, director of the Center for Bioethics at the University of Pennsylvania.

The coalition now says it hopes to use the failed bill to tip the balance in November’s close-run elections for Congress. “We’re going to do all we can to make this a campaign issue,” says Tony Mazzaschi of the Association of American Medical Colleges, a member of the CAMR. ■

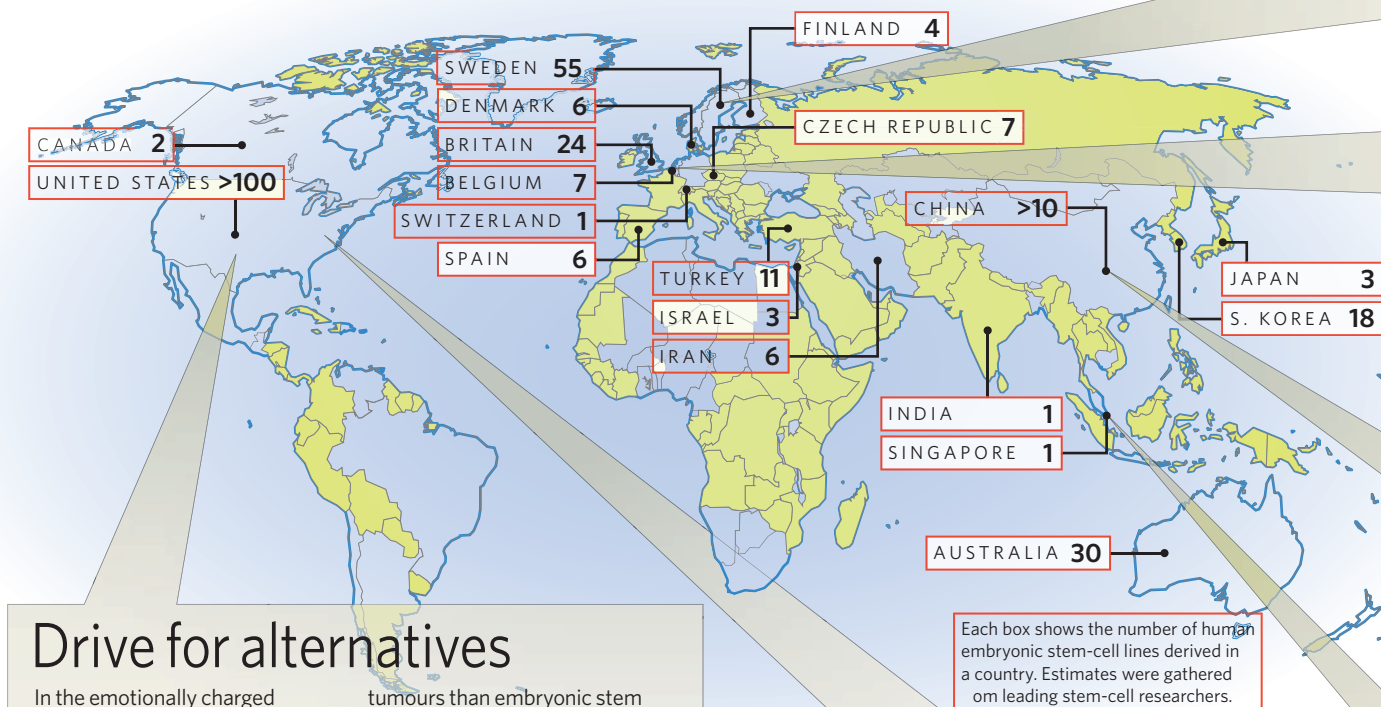
Meredith Wadman and Alison Abbott

THE COST OF RESTRICTIONS

Scientists argue that being unable to derive and work on new human embryonic stem-cell lines is a huge setback for medical research. With hundreds of such lines already in existence, why do researchers want so many, and what’s wrong with the old ones? See overleaf ▶

THE LURE OF STEM-CELL LINES

Nature investigates what human embryonic stem-cell lines have been derived worldwide so far, and why scientists are so desperate to work with new ones.



Drive for alternatives

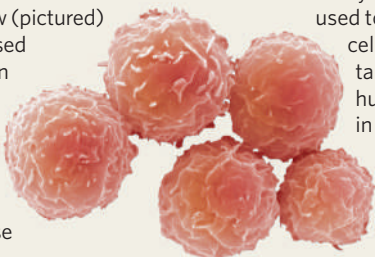
In the emotionally charged stem-cell debate, to be against embryonic stem-cell research is often portrayed as being for disease. So when opponents to loosening restrictions on US funding of embryonic stem-cell research took to the Senate floor last week, they were keen to emphasize their support for research that doesn't involve the destruction of embryos (see pages 329 and 335). Some senators exaggerated the promise of adult-derived stem cells, but they have real potential, as do other alternatives to embryonic stem cells.

Adult-derived stem cells are the only form of stem-cell therapy to make it to the clinic so far. For example, stem cells from bone marrow (pictured) have been used for more than 30 years to treat blood disorders. Adult stem cells are less likely to cause

tumours than embryonic stem cells, and less likely to be rejected by the immune system.

But adult stem cells are limited in other ways — most notably supply. Some adult stem cells can be grown in the lab, but not indefinitely. Isolating sufficient quantities of certain tissues, such as brain samples, could also pose a major technical hurdle.

A few scientists are taking a different tack and trying to harvest the potential of embryonic stem cells — particularly their immortality and flexibility — without destroying embryos to get them. For example, experiments in mice have shown that a single cell taken from a blastocyst, without destroying it, can be used to derive a cell line. But tampering with human embryos in this way may not address everyone's ethical concerns.



Diminishing returns

In August 2001, US President George W. Bush restricted federal funding to work on human embryonic stem-cell lines already in existence. Of 64 lines said to be available, some failed in culture and some were retracted by donors, leaving 21 approved lines. But why aren't scientists happy with those?

About half of the available lines grow slowly, making them virtually unusable, says Renee Reijo Pera of the University of California, San Francisco.

Another problem is that the cell lines have aged in the past five years and accumulated genetic mutations. There are frozen stocks, but many were created using protocols that

are outdated, meaning thawed cells can be difficult to culture.

Then there are the 'feeder cells' that form the matrix on which stem cells grow. All 21 lines were cultured using feeder cells from mice, so they may be too contaminated for use in humans.

There are also unexplained differences in how cell lines behave — for example, some differentiate more easily into certain cell types than others.

And researchers can't use disease-specific cell lines, which come from embryos with conditions such as cystic fibrosis: the Reproductive Genetics Institute in Chicago claims to have derived 36 such lines, but not one is approved for federally funded research.

Sweden steams ahead

At least 55 human embryonic stem-cell lines have been derived in Sweden: 30 at Cellartis in Gothenburg and 25 at the Karolinska Institute in Stockholm.

Do scientists really need so many lines to choose from? "Well, they're not all quite the same," says Outi Hovatta of the Karolinska Institute. "We've found that around 1,000 of the 30,000 genes we have checked in our cell lines are expressed differently." Each cell line

has characteristics of its own, she says.

All have the basic ability to grow in culture and to transform into different cell types. But some grow faster than others, and some are more stable in long-term culture. "Some differentiate more easily into particular tissues," adds Hovatta. "Many lines differentiate very nicely into nerve cells, for example, while others can turn into pancreatic cells, which are always rather tricky."

Europe's framework funding

A major research collaboration dedicated to human embryonic stem cells and funded by Europe's Sixth Framework Programme (FP6) kicks off next month.

The project, ESTOOLS, got €12 million (US\$15 million) from FP6, which has been matched by the collaboration, comprising 20 labs in 10 different countries. The project will study how stem cells differentiate into neurons, and will fully characterize 52 human lines.

Several FP6 programmes on

specific disorders and 11 new projects have just received ethical approval for work on human embryonic stem cells (pictured). And this week EU ministers agreed to continue funding in FP7.



V. STEIGER/SPL

Estimates from China

At least six laboratories in China are thought to have succeeded in deriving human embryonic stem-cell lines. Government funding for such research has significantly increased in the past five years, but there are no proper guidelines, and it is hard to monitor achievements. "There's a lot going on that you don't really know about," says

Jack Price of King's College London, who led a stem-cell team to China in 2004.

Estimates of the number of Chinese lines range from 10 to almost 70. Lingsong Li, who heads the Peking University Stem Cell Research Center, is cautious. He says that based on published evidence the most likely figure is "more than 10, but less than 15".

Clinical trials in sight

A big problem with the early human embryonic stem-cell lines is that they were not designed to be safe for use in humans. This week, Singapore-based biotech company ESI announced that it has derived four safe lines (produced in Australia) — with four more in the pipeline. It will make them available to researchers worldwide.

Regulatory agencies have not yet said what



standards cell lines would need to meet to be used in humans. But academic and biotech groups are gearing up to generate lines that satisfy Good Manufacturing Practice (GMP), a system for pharmaceutical products. Facilities that can derive and bank GMP lines are being built around the world.

It may also be possible to elevate existing lines to GMP status. The main problem with older lines is that they have been cultured with animal products. But Geron, based in Menlo Park, California, claims to have derived clinical-grade cells using two US-approved lines that were initially cultured using mouse cells. The company will apply for approval to start US clinical trials in 2007, using glial cells derived from human embryonic stem cells to treat spinal injuries.



Under siege: the weight of immigrants entering the United States via national parks is damaging fragile ecosystems.

M. YORK/AP

Wildlife caught in crossfire of US immigration battle

Richard Felger, a conservationist who runs the Drylands Institute in Tucson, Arizona, used to stop and share coffee with people he met crossing the desert along the US–Mexico border. Not any more. A couple of summers ago he was robbed at gunpoint in the Sierra Madres by some drug runners. Another time he was almost carjacked. “You can’t tell who’s friendly anymore,” he says.

Felger and other conservation biologists say that the biologically valuable lands along the border have become a war zone; increasing numbers of migrants and drug-traffickers are travelling north via the desert, which means more Border Patrol agents pursuing them in cars and helicopters. The dangers are making research into how all the activity is affecting the region’s endangered wildlife impossible. But researchers warn that proposals to erect hundreds of kilometres of fences and roads to try to stop the illegal crossings will damage many of the fragile species even further.

The US–Mexico border is one of the longest policed borders in the world, stretching some 3,200 km. Cutting through the Colorado river

delta, it then runs through the Sonoran Desert for much of its western length. The area is a conservation priority because its rich biodiversity is threatened by a declining water supply — humans use nearly all the water in the Colorado river. It hosts more plant species than any other desert, as well as vulnerable animals including the pronghorn antelope, and species of bat, owl and jaguar.

Over the past decade or so, tougher border enforcement at major crossings such as San Diego in California and El Paso in Texas, has squeezed migrants and drug runners out into the desert. That’s making life dangerous for local wildlife, and the biologists studying it.

“The Sonoran Desert system can be pretty fragile,” says Kathy Billings, chief ranger of the desert’s Organ Pipe Cactus National Monument park. Billings began her job not long after ranger Kris Eggle was shot and killed when on duty in a drug-related pursuit in 2002. Billings describes the illegal trails beaten into the desert by immigrants and smugglers. The desert is spangled with

trash — nappies, toothbrushes, backpacks and water bottles painted black for night walking. In places, “the smell of human waste is overwhelming”, she says. About a third of the park is now closed to visitors, and researchers must frequently be accompanied by armed guards.

“Today these groups are often carrying drugs and are a lot more dangerous,” adds Mark Briggs, a restoration ecologist who has worked with the conservation group the WWF, and Big Bend National Park in Texas. “I’ve had some fairly frightening experiences. But you go on with your work, and they go on with theirs.”

The dangers mean that data on how badly the environment is being damaged are scarce. “It’s a joke to say we can do pygmy-owl surveys with the amount of activity you see down there,” says Jenny Neeley, southwest representative of Defenders of Wildlife. Neely, based in Phoenix, Arizona, is co-author of a report about how illegal immigrants and the Border Patrol hurt wildlife.

Peter Morrison, director of the Pacific

“The smell of human waste is overwhelming.”

Biodiversity Institute in Winthrop, Washington, has been able to do some research on the problem. In a survey of the Buenos Aires National Wildlife Refuge, for example, he recorded an average of 4 km of illegal trails per square kilometre of land. Half the trails were found in the habitat of the endangered Pima pineapple cactus. Morrison presented his results last month at the annual meeting of the Society for Conservation Biology in San Jose, California. But he and his colleagues are now avoiding certain areas. "It's getting a bit scary," he says. "A lot of people in the field are questioning whether they want to keep doing this work."

This election season, US politicians have focused their attention on the problems at the border. In the past couple of months President George W. Bush has sent thousands of National Guard troops to aid the 9,000-strong Border Patrol. And duelling bills in the House and Senate propose hundreds of kilometres of extra fencing and illuminated roads. "We are going to see some fencing built, and some roads built," says Marshal Fitz, director of advocacy for the American Immigration Lawyers Association based in Washington DC. "It's just a matter of how much and when."

But although conservationists are pleased that attention is being paid to the situation, they are horrified by the influx of troops and the proposed fences. "What they are putting up is like a Berlin wall, and the border-control guys are creating roads left and right," says Briggs. "It has been a huge setback for the wildlife."

"It's very important to try to maintain a

"I don't see anyone addressing why immigrants are coming in the first place."

wildlife-friendly border," agrees Morrison. "If we put up a 12- or 25-foot high wall, it stops everything except things that fly from crossing. That would be a real disaster in the long term."

Researchers are particularly worried about the easily spooked migratory Sonoran pronghorn antelope. Fences would fragment the endangered antelope populations into even smaller groups, and the extra stress caused by border-control traffic might push them over the brink. Meanwhile trash piles attract ravens, which might eat threatened tortoises, and all-night lights disrupt bugs, bats and nocturnal cats such as the ocelot and jaguarundi.

There is an alternative to solid fences. Last month, Organ Pipe installed a steel rail, one metre off the ground, which stops cars and trucks yet allows wildlife through. Park biologist Tim Tibbets says that since the barrier, funded by the National Park Service, was erected, vehicle traffic has virtually stopped. Other parks and the nearby Tohono Oodham Indian reservation are looking into putting up similar barriers. The Border Patrol is experimenting with vehicle barriers and is currently coordinating a 10-km section, with a view towards extending that to 133 km, along the Arizona border.

But there seems to be more political support for the heavy-duty barriers. And many argue that fences alone won't solve the problem — in terms of immigration or conservation. "I think there is a heavy militaristic hand in the US response to this," says Briggs. "There's momentum now towards putting in these large fences... I don't see anyone addressing why immigrants are coming in the first place."

"No matter what kind of wall we put up we are going to have problems," says Morrison. "The fundamental solution will be to reduce the incredible economic disparity between the countries."

In the meantime, Neeley and Defenders of Wildlife contend that the Border Patrol has neglected the environmental impacts of its work, and has not fulfilled its legal responsibilities under the Endangered Species Act.

In a statement to *Nature* the Border Patrol said: "Our primary responsibility is to prevent terrorists and their weapons from entering the country... We will continue to move forward and work with other federal agencies to minimize our effects on the environment." But in an informal conversation, a Border Patrol spokesman said he felt that conservation projects are just not on the agency's radar screen.

"There are owls on the southwest border?" he added.

Emma Marris



The Sonoran Desert, a highway for drug traffickers, is also home to the endangered northern pygmy owl.

ON THE RECORD

"Common sense tells you that you want professionals trained to the Nth degree to do jobs like astronaut jobs."

Astronaut Jerry Linenger disapproves of plans by a private company to offer spacewalks to the public — at a total cost of \$35 million.

"Danger! Water contains high levels of hydrogen."

Officials in Louisville, Kentucky, post signs to discourage residents from playing in the city's fountains.

Sources: *Washington Post*, *Louisville Courier Journal*

SCORECARD



Long-range forecasts
Japanese scientists unveil plans to create a 30-year weather forecast using their Earth Simulator supercomputer.



Plants in space
China announces that it will launch a satellite this autumn solely to breed seeds in microgravity.



Australian science
A government audit shows that Australia may face a shortage of 20,000 scientists and engineers by 2012.

NUMBER CRUNCH

The Union of Concerned Scientists recently surveyed 997 researchers at the US Food and Drug Administration (FDA) and found:

70% believe the agency lacks the resources to protect public health.

60% knew of commercial interference in the FDA process.

18% had been asked to inappropriately exclude or alter technical information or their conclusions.

40% said they could not publicly express concerns about public health without fear of retaliation.

Source: *Union of Concerned Scientists*

Carbon credits for the Joneses

It sounds like a triumph for the doctrine that people should think globally but act locally — and like a nightmare scenario for libertarian opponents of big government. Last week, UK environment secretary David Miliband suggested issuing all British adults with an annual carbon allowance. Advocates say the system is fair and would focus people's attention on conserving energy. But could it ever succeed?

Here is how the system would work. For transactions that involve direct purchases of energy, such as buying petrol or paying fuel bills, a person would hand over money and some of the carbon credits he or she had been allocated by government. If those credits ran out, the person would have to buy extra when paying for the fuel or electricity. By regulating the amount of personal credits handed out each year, the government could cap total carbon emissions and help tackle climate change.

"Instead of banning particular products, services or activities — or taxing them heavily

— a personal carbon allowance enables citizens to make trade-offs," said Miliband as he floated the idea on 19 July.

Civil servants will look into the proposal and report back to government next year. Researchers who have studied the idea say domestic quotas are a sensible way to extend emissions trading to the personal level — such trading is already used to limit emissions from some European industries. And unlike a blanket carbon tax, it encourages individuals to think about their emissions. "It makes carbon more visible," says Richard Starkey, a carbon-policy expert at the University of Manchester, UK.

Starkey also points out that low-income families, which tend to use less energy than higher earners, could save and then sell carbon credits. At current UK emission levels, he reckons each individual would receive around 1.25 tonnes of carbon. For an idea of scale, a 200-km journey in

a car that uses petrol would use about 1% of that. The allowance would be worth only a few tens of pounds (or US dollars) at today's prices, but if policies are enacted to meet the ambitious UK target of reducing emissions to 60% below 1990 levels by 2050, that would rise significantly.

Such ideas are going down rather less well in the United States, where environmental activists privately hope the UK proposal will attract as little attention as possible. Oil-industry groups have previously claimed that attempts to tackle climate change such as the Kyoto Protocol will damage lifestyles — an accusation that most politicians, particularly in the United States, are careful to avoid. Personal carbon credits may sound too much like energy rationing. "It would be awfully hard to explain this to Americans," says one senior US environmental advocate, who asked not to be named. ■

Jim Giles

"It would be awfully hard to explain this to Americans."

Medical association acts to ensure journal autonomy

A row over editorial independence at the *Canadian Medical Association Journal* (CMAJ) has ended with the association agreeing to take control of the journal from its for-profit subsidiary.

Two CMAJ editors were sacked in February after executives at the subsidiary, CMA Holdings, intervened in a story about women's access to the morning-after pill (see *Nature* 440, 10–11; 2006). Most of the editorial board resigned in protest.

The journal will now be run by the association itself, which has promised to make “editorial integrity” an explicit goal. The association will also be informed in advance of any controversial content, but will not have the power to change or cut anything. The new rules are the result of a review by an eight-member panel of academics, journal editors and representatives of the CMA.

Founder of gene therapy convicted of child abuse

William French Anderson, often known as the ‘father of gene therapy’, was convicted on 19 July of repeatedly sexually abusing a colleague's young daughter. Anderson, 69, could be given up to 22 years in prison when he is sentenced in November.

The University of Southern California in Los Angeles, where Anderson directs the Gene Therapy Laboratories, immediately began proceedings to fire the physician, who is famed for initiating breakthrough clinical trials at the National Institutes of Health. Anderson was jailed after the verdict and is undergoing psychiatric examination. His attorney says he will appeal.

In 1997, when the victim was ten years old, Anderson began a five-year-period of abuse (see *Nature* 430, 600–601; 2004). Crucial trial evidence included surreptitious audio tapes that the police made in 2004 of Anderson acknowledging the activity to his victim.

Fifteen-year goal to transform rice yields

Can rice be bred or engineered to photosynthesize as efficiently as corn (maize)? Researchers at an International Rice Research Institute (IRRI) workshop say that it can, and have identified two strategies for achieving this ambitious goal.

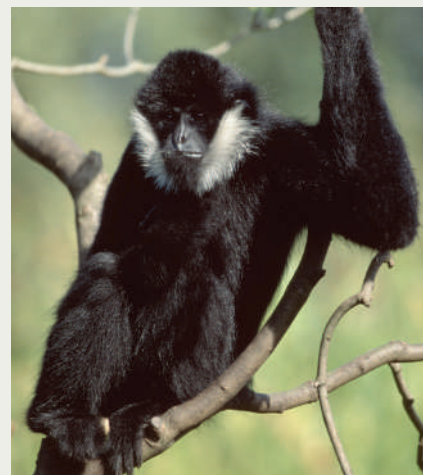
Maize uses carbon dioxide more efficiently in photosynthesis, so transferring the underlying biochemical machinery to rice would lead to far bigger yields. At the

Gibbon has cheek to join the genome sequencing elite

The Northern white-cheeked crested gibbon (*Nomascus leucogenys leucogenys*, pictured) is the latest species to be lined up for genome sequencing, the National Human Genome Research Institute, based in Bethesda, Maryland, announced on 19 July.

The primate is of interest in part because it has an unusually large number of segmental duplications — large, almost identical stretches of DNA that are present in two places in its genome. Similar duplications are known to be associated with disease and birth defects in humans.

Also scheduled for detailed sequencing are the domestic cat (*Felis catus*), the African savannah elephant (*Loxodonta africana*), five species of fungus that infect humans and up to 50 strains of the yeast *Saccharomyces cerevisiae*.



WEGNER/ARCO/NATURE PL

IRRI meeting, held on 17–21 July in Laguna in the Philippines, a consortium of about 20 research institutes set the goal of achieving this within 15 years.

The groups will try to introduce the more efficient photosynthetic apparatus by studying wild relatives of rice that may have maize-like photosynthesis, or by introducing genes from maize into rice. The aim is to meet growing demand for food in developing countries.



Rice may benefit from receiving genes that encode the photosynthetic apparatus of maize.

Nobel laureate accused of blocking MIT appointment

The Massachusetts Institute of Technology (MIT) has launched a review of its neuroscience research following allegations that a Nobel-prizewinning neuroscientist used his position to interfere with the appointment of a researcher at MIT.

In a 15 July story in *The Boston Globe*, Susumu Tonegawa, head of MIT's Picower Institute for Learning and Memory, was accused of deterring Alla Karpova from accepting a position at the McGovern Institute for Brain Research, which is based in the same building as the Picower. Karpova has since rejected MIT's offer and accepted a job at the Howard Hughes Medical Institute's Janelia Farm research campus in Virginia. The allegations have been disputed

by some of Tonegawa's colleagues.

According to *The Boston Globe*, MIT president Susan Hockfield acknowledged that there are “ongoing tensions among MIT's neuroscience entities”.

The review, which is due to report within the next few months, will examine the structure of MIT's neuroscience research and how to boost cooperation between groups.

See related blog entry on Nature Network Boston: <http://network.nature.com/boston/community/view/35>

Private funds set to counter ‘crazy’ retirement rules

Germany's much-criticized retirement regulations, under which university staff are compelled to leave when they reach 65, are to be tackled by a private funding agency.

The Frankfurt-based Hertie Foundation, which funds neuroscience projects, said on 16 July that it will provide a series of grants for neuroscientists over 60. The scheme is designed to prevent the loss of some of the country's most experienced scientists to the United States.

Federal research minister Annette Schavan supported the programme at its launch, saying that the German rules for researchers were “crazy and unsuitable for the future”. The foundation acknowledged that only a few older scientists remain active and successful, and said that its selection criteria were stiff.

The first recipient is 63-year-old neurologist Thomas Brand from the University of Munich.

Correction

Our “Science at the Solstice” News Feature (*Nature* 441, 1040–1045; 2006) should have reported that Diana James Junau works with Grand Perfect Conservation, and the picture of her shows her working in its laboratory, not at a university facility.

BUSINESS

Break with tradition

Traditional medicine has spent decades in the wings of pharmacology. Now India is pushing it to centre stage, as **K. S. Jayaraman** reports.

For years, the drug industry has been curious about traditional medicine — especially the venerable systems of India and China. Now, the Indian government has taken a step that could open the way for greater commercial exploitation of its traditions around the world.

In the past few years, India has developed a huge electronic database known as the Traditional Knowledge Digital Library. Late last month, the Indian cabinet agreed to give patent offices around the world access to the library, to make sure that patents are not granted on existing Indian remedies. And the government may soon go one stage further, inviting major international drug companies to collaborate with Indian researchers on deriving drug candidates from the library's contents. It hopes to boost the country's public health care in the process.

But the move to share the library's content has sharply divided opinion in India, where the knowledge is seen as an important part of the country's cultural and intellectual heritage.

Advocates of sharing say that the database, which has been under construction at the National Institute of Science Communication and Information Resources in New Delhi since 2000, could have a major impact on the process of drug discovery. The database has the potential to "slash the cost of drug development", says Vinod Gupta, a computer scientist and director of the institute. "We have a treasure chest of plant-based medicines, created by experimenting directly on man for hundreds of years."

Others are not so sure. They worry that India risks losing out by sharing its knowledge with outsiders. Purveyors of traditional medicine fear that international companies will grab control of the information. "It is hard to believe that the multinational drug companies are interested in collaborating on traditional-medicine research in order to promote it," says P. Ram Manohar, research director of Aryavaidya Pharmacy, which produces drugs based on traditional knowledge in Coimbatore. "Their interest would be confined to using it to develop new drugs — over which they could exercise

control." And that would be of little help to India's healthcare agenda.

Some also doubt that the information will really yield the blockbuster drugs that architects of the database are hoping for. Two major drug companies that are active in India — Pfizer and Merck — declined to respond when asked whether the database was of interest to them.

Medicine bags

Traditional Indian medicine consists of three main systems, known as *Ayurveda*, *Siddha* and *Unani*. Between them they use about 1,500 medicinal plants, a third of which are used in drug formulations. The Traditional Knowledge Digital Library already contains information on 145,000 formulations, with another 15,000 due to be added in the next year.

The creation of the database reflects India's gradual acceptance of international patent law. For decades, the country had largely ignored foreign protections, and allowed its drug companies to sell generic copies of patented treatments. But in 1995 it joined the World Trade Organization's intellectual-property agreement.

At that point, India started to use foreign courts to tackle abuses of its traditional knowledge. In 1997, for example, it won a landmark case in the United States, which overturned patents on the use of turmeric to treat wounds. Three years later, it got the European Union to revoke a patent on the use of an antifungal treatment derived from the Indian neem tree.

In an effort to mount a more systematic defence of its traditional knowledge, the Indian government decided

to document as much of it as possible in a digital format. And on 29 June, the cabinet agreed to make it available, in confidence, to patent examiners abroad, so that they can search the database before granting patents on any of its components.

But critics have questioned whether adequate safeguards will be in place to prevent unauthorized access to the library. The best



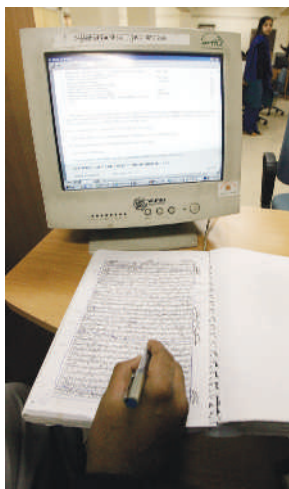
Power plants: India is working out how best to exploit its traditional herbal lore.

safeguard, according to some Indian scientists, would be a programme to actively patent the contents of the database. According to environmental activist Devinder Sharma, China has filed thousands of domestic patents on its traditional medicine. Indian observers also say that China has done serious research to build up that knowledge — and has promoted it effectively in global markets.

India's approach will be quite different, according to Gupta. Although no detailed plans have been drawn up yet, it intends to open its database to drug firms — both Indian and foreign — on commercial terms. "This is our next step and that is really where the country is going to benefit," he says, adding that "informal" discussions with multinational drug companies indicate they are keenly interested. But the proposal has yet to be cleared with the government, he says. "All this will take a few months."

Local heroes

Gupta thinks the approach will benefit the country because multinational firms, who are unfamiliar with Indian traditional medicine, will need to collaborate with local companies and research institutions to make use of the database. The strategy is backed by Ranjit Chaudhury, chairman of the Indian Clinical Epidemiology Network Trust International in New Delhi, who adds that so far India's record of exploiting traditional knowledge is nothing



India's database of traditional medicine has 145,000 formulas.



SOUTHINDIAPICTURE/ALAMY

to write home about. “We have not been able to develop any drugs ourselves,” he says, noting just one exception — guggulipid, a compound extracted from the myrrh tree, and approved domestically as an anticholesterol drug. “So why not give Western companies a chance?”

That sanguine outlook is rejected by Suman Sahai of Gene Campaign, a pressure group based in New Delhi that strongly opposes the opening of the database to commercial firms. Foreign companies won't need local partners, she says: “they can follow the leads on their own”. And Nitya Anand, a former director of the Central Drug Research Institute in Lucknow, warns that even if the major firms collaborate with local researchers, India may not be able to guide them to pursue agendas relevant to pressing local healthcare needs.

Other critics say that Western companies won't be able to do much with the traditional remedies because they are obsessed with single chemical entities. “The single-molecule model does not apply to our medicines, which work because of the synergistic effect of several molecules,” says G. Satyavati, former chief of the Indian Council of Medical Research.

But many researchers in India take a brighter view. “Using traditional knowledge to discover medications is an idea whose time has come,” says Vis Niranjan, president of RxMD Private, a management consultancy based in Chennai. “Cures for many illnesses are found in nature. The information contained in the library is phenomenal — and it is time someone mined the data to identify cures.” ■

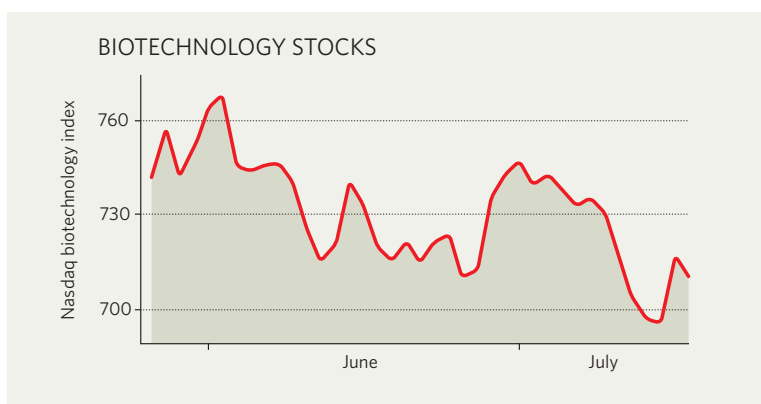
IN BRIEF

VACCINE VENTURE Swiss drugmaker Novartis has announced plans to build a \$600-million, state-of-the-art production plant for flu vaccine in Holly Springs, North Carolina. The plant — more than a third of which will be paid for by the US government — will be the first in the United States to derive vaccines from cell culture rather than the chicken eggs commonly used at present. The company says its facility is designed to produce 50 million doses of seasonal flu vaccine annually, and up to 150 million doses of avian-flu vaccine if required.

CHINA CRISIS Amnesty International, the human-rights watchdog, has accused Google, Yahoo and Microsoft of contributing to ‘Internet repression’ in China by cooperating with the country's authorities. “The apparatus of Internet repression is considered to be more advanced in China than in any other country and companies are particularly willing to cooperate with the Chinese government,” Amnesty says in a report issued on 20 July. Yahoo has faced a consumer backlash in the West, after giving the police the identities of two dissident Chinese writers, who are now in prison.

GREEN FOCUS The Ford motor company has said that it will spend £1 billion (US\$1.9 billion) over six years in Britain on research and development into cleaner engines. The company says that 9,500 engineers will be deployed in the effort. It intends to create a version of its most popular car — the Ford Focus — that delivers 70 miles per gallon. The announcement has been welcomed by the government, but unions note that it involves the redeployment of existing resources, not fresh investment.

MARKET WATCH



This week, Wood Mackenzie, an Edinburgh-based research and consulting firm, reviews recent trends in biotechnology stocks.

Biotech continues to retreat from its high point in February, although the rate of decline has slowed: the Nasdaq biotechnology index is down 4% over the past eight weeks, and 12% since the start of the year. Broader indices are also falling in a volatile market.

Amgen of Thousand Oaks, California, has fared particularly badly, falling 5% over the past eight weeks and 20% so far in 2006. Investors believe there is a growing threat to Amgen's erythropoietin drugs for treating anaemia, which generated \$5.8 billion in sales in 2005 — nearly half of total turnover. Rival Roche of Basel, Switzerland, has a second-generation erythropoietin drug, called CERA, which is likely to reach the market in 2007. And the European Union has cleared

a path for the approval of generic versions of some biological drugs, including erythropoietin.

Amgen is given more ‘weight’ in the index than any other company, so its losses are an important factor in the overall drop. But many other listed firms have suffered.

Shares in Anadys Pharmaceuticals of San Diego, California, lost two-thirds of their value after the company suspended a phase I trial of its hepatitis-C treatment and its chief executive announced his forthcoming departure. Stock prices in another San Diego company, Neurocrine Biosciences, dropped by three-quarters after problems with its insomnia drug candidate, indiplon.

In a period of general market anxiety, biotech shares are particularly vulnerable to bad news. Now, strong second-quarter results will be needed to bolster confidence in the sector. ■



THE TROUBLE WITH REPLICATION

The idea that readers should be able to replicate published scientific results is seen as the bedrock of modern science. But what if replication proves difficult or impossible? **Jim Giles** tracks the fate of one group of papers.

There is nothing particularly unusual about the 6,893rd issue of *Nature*. Published four years ago, it covers the usual mix of disciplines. One paper describes a development in quantum computing¹, another contains the genome sequence of a slime mould². The subsequent impact of the papers has also been within the normal range: a handful have been referenced hundreds of times each, but most have notched up only a few tens of citations.

A third feature of the edition, though just as normal, is more surprising: at least two of the papers we published on 4 July 2002 contain results that may not be replicable. There is nothing suspicious about the papers, nor any suggestion that their authors are anything other than excellent scientists. Nor was that week particularly odd, and there is no reason to think that other journals publish fewer problematic papers. It is simply the case that the replication of results, a process absolutely central to science, is not always possible.

"If you want to know whether a duck is

crossing the street, you look twice," says Harry Collins, a social scientist at Cardiff University, UK, who cheerfully describes himself as the world expert on replication. Replication in science, he says, is the same: it is a way of being sure that something really exists, and the process by which tentative discoveries acquire textbook status. If, on the other hand, attempts to replicate a result meet failure again and again, that result will end up being discounted.

But, as Collins would be the first to point out, the situation is more complicated than that. For a start, many papers, especially in minor journals, go unreplicated simply because of lack of interest (a third of all papers are never again cited in the scientific literature). Their replicability thus becomes moot.

More concerning for scientific progress is what happens when attempts to replicate an interesting experiment fail. At what point does 'unreplicated' give way to 'unreplicable'? And what is the best way to find out where on this scale a particular result might sit — by gossiping in the bar? By reading another paper? Or

by some semi-formal mechanism in between?

To see replication and its absence in practice, and to ask whether journals and scientists are doing enough to monitor it, I tracked the fate of the 19 papers in issue 6893, asking their authors and others in the field whether the results had been reproduced. In a large majority of cases they had. For example, the procedure with which Ron McKay and his colleagues at the National Institute of Neurological Disorders and Stroke in Bethesda, Maryland, treated rats suffering from an analogue of Parkinson's disease with embryonic stem cells³ has since been repeated in rats elsewhere and in other animals. Papers on the formation of silica films⁴ and on yeast genetics⁵ have quietly racked up citations as the results have been replicated and built upon.

In other disciplines, results are corroborated rather than reproduced. If an identical description of the fossil found by palaeontologist Jennifer Clack⁶, based at the University of Cambridge, UK, was published elsewhere, for example, it would look more like plagiarism

than replication. Her interpretation, though, has undergone something like replication; similar fossils that date from the same period have since been found and described in a way that conforms with her conclusions. In the case of the genome of *Dictyostelium discoideum*², an amoeba, few researchers would see the need to repeat the sequencing from scratch; in any case, the genome stored on the dictybase.org website can be updated should errors be identified.

Giant's signature

But for other papers from the 4 July issue, textbook status looks a long way off. One of those was authored by Sean Brittain and Terrance Rettig, both then based at the University of Notre Dame in Indiana. Their finding was an exciting one: they claimed, for the first time, to have seen H_3^+ ions in the disk of gas and dust surrounding a young star⁷. H_3^+ is seen in the atmospheres of Jupiter and Saturn, suggesting that the astronomers had spotted a gas giant in the act of formation.

Yet right from the start, other researchers wondered whether Brittain and Rettig really had seen H_3^+ . The evidence was in the form of distinct frequencies of infrared radiation: H_3^+ emits at three particular frequencies, and Brittain and Rettig reported detecting emissions in only two of those three. Takeshi Oka of the University of Chicago in Illinois wrote a cautiously optimistic News and Views commentary on the finding in the same issue⁸ but had his doubts about the result. "We used our earliest observation time to check," he says now. "We couldn't see it."

Over the next year, the two sets of authors exchanged their raw data in a bid to resolve their contradictory results. Such exchanges are never easy, given that the scientists are to some extent putting their reputations on the line. In this case, neither side seems to have completely accepted the other's conclusions. But Oka has since published a paper describing how he failed to find any of the three H_3^+ lines⁹. Rettig says that he and Brittain no longer "promote our very tentative interpretation that the unidentified lines might be H_3^+ ". But a second key result from the original paper — the detection of carbon monoxide in the same disk — has since been confirmed.

Attempts to replicate another result from issue 6893 — one that, like the H_3^+ , was considered worthy of mention on the cover — have created far more controversy. In July 2002, stem-cell biology was national news in the United States. Because some stem cells are pluripotent — they can develop into many different cell types — they offer a chance of replacing the tissues lost to old age or disease. But for opponents, most notably those on the influential religious right, the fact that cells for research are extracted from human embryos renders the process unethical, whatever its promise (see pages 486 and 491).

As the two sides battled over plans to regulate the field, a team from the University of



Minnesota pitched in with a paper that seemed to offer a peaceful solution. Catherine Verfaillie and her colleagues described how they had isolated fully functioning stem cells from adult human bone marrow¹⁰. If the results were correct, all the benefits of stem cells could be realized by taking samples from the patient involved — no embryos, no cloning. To those with moral objections this sounded vastly preferable; to others it simply sounded easier. It looked like a win-win situation.

But four years later, the implications of the paper are still far from clear. "People found the paper amazing," says Stuart Orkin, a stem-cell biologist at Harvard University. "But there has been very little published literature since. There

"If you want to know whether a duck is crossing the street, you look twice"
— Harry Collins



has been no clarification of what those cells are."

The story will be a familiar one to many biologists. After publication, rival labs fell over themselves to reproduce the results. Many contacted Verfaillie with requests for the cells and the reagents used to create them, or to ask for more details of the experimental protocol involved. Yet progress was not smooth. Several high-profile groups sent researchers to Minnesota to learn how to extract and culture the cells, and then brought the result into doubt by stating that they could not repeat the process back in their own labs.

The test of time: most papers in this 2002 issue are looking good for their age — but who could tell which would stand up to testing?



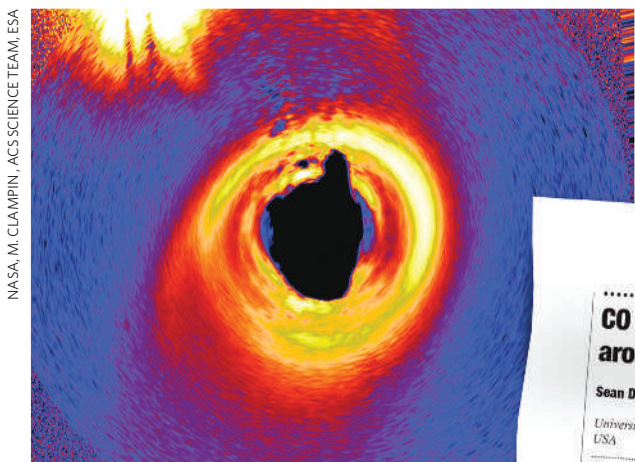
Verfaillie counters that the procedure takes up to six weeks to master and that those who stayed for long enough have cracked it. Some, indeed, have published results¹¹.

Shy journals

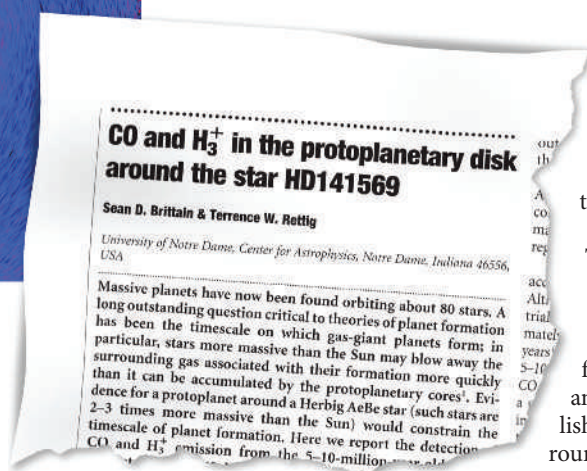
Verfaillie adds that her own team has since ironed out problems with the serum it used and will soon publish a comprehensive methods paper describing the new protocol. But researchers who think they have derived pluripotent stem cells from human bone marrow using related but not identical techniques to Verfaillie's, including Dominique Bonnet at Cancer Research UK's Lincoln's Inn Fields Laboratories in London, have still found it difficult to get their results into print. They say that referees for top journals, aware of the difficulties independent groups have had in replicating Verfaillie's results, are now so sceptical that it is hard to publish their findings.

Verfaillie's protocol maybe indeed be unusually complicated, but it illustrates a common problem: it is often hard to tell whether an inability to replicate a result is due to a group's failings or a flaw in the original paper. The reason is often the countless tiny details of experimental method that are omitted from the methods sections of papers but can influence results. "Things are different in different labs for very subtle reasons," says Gillian Murphy, a cell biologist at the University of Cambridge, UK. "The water can be different. We're about to move labs, and my group is very concerned that delicate cells might hate something in the new pipes."

S. GELDER



Now you see it: one paper (below) reported a tantalizing hint that gas giants might be forming in the dust and gas around a young star (left).



This issue is not confined to the biological sciences, as Collins's research reveals. In his 1985 book *Changing Order*¹², he quotes a physicist on the difficulties of recreating an exact copy of a piece of experimental kit: "It's very difficult to make a carbon copy. You can make a near one, but if it turns out that what is critical is the way he glued his transducers, and he forgets to tell you that the technician always puts a copy of *Physical Review* on top of them for weight, well, it could make all the difference."

A paper can never be a foolproof recipe for the replication of its results because this sort of information, which the chemist and philosopher Michael Polanyi called "tacit knowledge", can never be entirely captured in a scientific paper. It is thus not in principle possible to tell whether a failure to replicate is down to a lack of this tacit knowledge or a flaw in the original result. In practice, researchers compensate by exchanging tips by e-mail and at conferences. Replication is a social phenomenon, which accounts for the interest of sociologists such as Collins. But because the social interactions are not recorded anywhere, it is hard to consult or build on them.

Fraud and its fallout

This becomes a particularly vexed problem in cases of fraud. Just a few weeks before issue 6893 was published, a scandal hit nanotechnology. Jan Hendrik Schön of Bell Labs in Murray Hill, New Jersey, was one of the brightest prospects in the field, his string of high-profile papers in top journals a remarkable feat for a researcher in his early thirties. But in May 2002 Schön's world began to unravel: people noticed that data in some of his papers had been manipulated. Later that year he was fired, with many of his papers withdrawn as fraudulent.

Allen Goldman's lab at the University of Minnesota in Minneapolis was one of many that wasted much time trying to replicate Schön's work — specifically his findings on the superconductivity of spherical carbon molecules known as buckyballs. A postdoc and two graduate students spent more than a year trying to make the things Schön had described in the papers happen anew in their own lab,

convinced that they were failing because they could not replicate Schön's procedures. Others recount similar experiences: "A postdoc of mine burned up a couple of years of his life," says Robert Dynes of the University of California, Berkeley.

The experience was not all bad: Goldman notes that the work they did inadvertently led his group towards more discoveries. He also says he is now far more critical about the papers he reads in journals. But while his team was trying and failing, others were having

"Rumours go round when part of a study doesn't work in other people's hands"
— Chris Surridge



similarly frustrating experiences — experiences that were only discussed at meetings and during the odd phone call between friends in rival labs. Had something appeared in the peer-reviewed literature, Goldman says he would probably have realized more quickly that something was wrong.

One obvious solution is for journals to publish more papers that describe failures to replicate results. Most top publications have procedures for dealing with data that conflict with previously published results, although they demand that authors amass a comprehensive data set before allowing them to question a published result. *Nature* does this through Brief Communications, and like most journals publishes only a handful a month.

Yet few scientists contacted by *Nature* suggested that the journals expand this activity, or

lower the thresholds they require for questioning a result. "Most failures to replicate exhibit incompetence," says Collins, whose feelings sum up those of many scientists. "It would be misleading to publish each one." At *Cell*, editor Emilie Marcus says she would be reluctant to publish a statement saying that someone had simply failed to replicate a paper. "Thorough attempts to reproduce a result should be published," she says, "If you want to claim publicly that someone is wrong, that takes a certain degree of evidence."

This leaves editors in a dilemma. The findings in papers are often hard to reproduce. Readers want to know if they should bother trying or instead dismiss the results as flawed. But the data that could help answer that question often lie unpublished in lab notebooks. "Rumours go round when part of a study doesn't work in other people's hands," says Chris Surridge, an editor at the open-access publisher the Public Library of Science in Cambridge, UK. "But it's very difficult to get the information into the public domain."

Electronic coffee

Scientific publishing's move online may be a big help here, opening up new possibilities for sharing and commenting on papers and methods. If these can be harnessed, coffee-break conversations about replication could be shared with the entire scientific community, allowing problems to be cleared up more easily and frauds discovered more quickly.

The most obvious first step towards this is simply to allow comments to be posted on published papers. Several publishers, including the open-access journals hosted by BioMed Central, already have this sort of comment function but have found that it is not widely used. That could simply be because the facilities need promoting, as BioMed Central publishers say they are now trying to do. Another reason could be that only the most successful researchers are confident enough to criticize others in this public way, even if journals allow it.

PLoS ONE, a new online-only journal from the Public Library of Science, will take the comment model further than anyone else when it launches later this year, with various functions for promoting discussion. Users will, for example, be able to annotate papers, including methods sections, with their own comments.

In a bid to tackle information overload, comments — or perhaps the people making them — will be rated by visitors to the site. Users will then be able to see only comments above a certain rating — a system pioneered by Slashdot.com, a technology website. Authors will also be able to correct mistakes or misleading statements, and others can link to improved methods published elsewhere. The aim, says

Surridge, who is managing *PLoS ONE* during its launch, is to recreate the kind of discussion that takes place in front of conference posters. To help keep things informal, the publication is considering allowing commenters to use nicknames — but only if they provide the site with academic bona fides.

This is not the only way in which the new journal hopes to shake things up. *PLoS ONE* also has an unusual publication policy: it will not consider the novelty of a result when deciding whether to publish a paper. As long as a result is judged by referees to be methodologically sound, and the author can provide the open-access publication fee, it will be published. That will, says Surridge, make it easier to publish papers that cast doubt on previous results, as well as those that confirm them. Sceptical editors at other journals say it will be hard to attract submissions to a journal that sets such a low bar for acceptance. Referees might, for the same reason, also be reluctant to review papers for the journal.

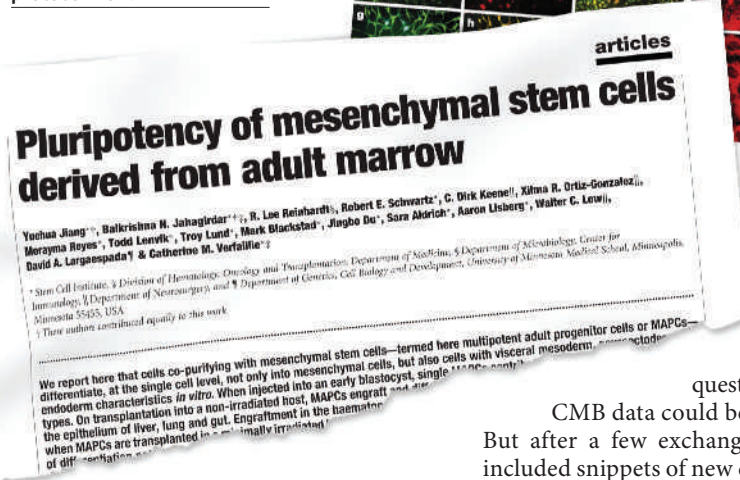
Here's how

Another possibility is to pay more attention to the methods sections of papers. The fact that publishers see this as a potentially important market can be judged by the launch in June 2006 of two journals devoted exclusively to publishing experimental protocols. By giving methods sections the same editorial care as a full scientific paper, from peer reviewing to archiving in a dedicated and searchable website, the journals should allow authors to detail the often critical minutiae of their experimental methods. "When I started in science I was told that I should be able to repeat an experiment by reading the paper, but that is almost never the case," says Michael Ronemus, editor of *Cold Spring Harbor Protocols*, published by Cold Spring Harbor Laboratory Press. "Editors tend to ignore supplementary information. It doesn't get the same scrutiny."

At the other new journal — *Nature Protocols* — editors say they will encourage authors to include a troubleshooting section. Both publications will include discussion facilities that will allow researchers to talk to the original authors and other users of the protocol. These could prove ideal places to resolve problems that crop up during attempts to reproduce a previous result. "The forums will go a long way to resolving conflicts," says Ronemus.

Another possibility is to link to conversation happening elsewhere on the web. ArXiv, a store of online physics preprints now hosted by Cornell University, has from its inception been an online trendsetter. Last August it implemented a 'trackback' function: a system that allows online discussions about a web page, in this case a preprint, to be easily linked to the original page. Following trackbacks from arXiv papers typically leads to physicists' blog posts, in effect

A hard act to follow: this paper said its authors had persuaded bone-marrow cells to turn into different types of body cell (right). Others failed to make the protocol work.



opening up a web of discussion that is something between open peer review and a coffee room at a conference. This might be disturbing for authors, who can now see their papers dissected in public, but it is a great way into the community's views on a paper and offers the benefits of informality and even anonymity.

"Junior people are often reluctant to have their name attached to negative comments," says Jacques Distler, a string theorist at the University of Texas, Austin, who helped to

"Junior people are often reluctant to have their name attached to negative comments"
— Jacques Distler



develop the trackback function at arXiv. "They don't know if it's someone they are going to be relying on later for a job."

Many links from arXiv take users to CosmoCoffee, a forum for discussion of cosmology and high-energy physics in general, and arXiv papers in particular. "Our motivation was to make papers more accessible, but it would be absolutely fantastic if it could resolve controversies as well," says Sarah Bridle, an astronomer at University College London and a co-founder of CosmoCoffee.

In October 2004, for example, an intriguing paper appeared in which the authors claimed to have set limits on the mass of the neutrino using data on the cosmic microwave background (CMB)¹³, a faint glow of radiation left over from the Big Bang. The paper raised some eyebrows on CosmoCoffee, as users

questioned whether the CMB data could be used in this way. But after a few exchanges, one of which included snippets of new data that had been generated to test out the conclusion of the preprint, users decided that the result made sense. For physicists wondering whether a particular result is robust enough to build upon, such discussion could prove a powerful resource. It is also, unavoidably, open to malicious attempts to undermine a particular researcher.

Forty years ago, the Nobel-prizewinning immunologist Peter Medawar declared that all scientific papers were frauds, inasmuch as they describe research as a smooth transition from hypothesis through experiments to conclusions, when the truth is always messier than that. Comments, blogs and trackbacks, by expanding the published realm beyond the limits of the traditional paper, may make the scientific literature a little less fraudulent in Medawar's sense, and in the more general one. They could also help the many frustrated scientists struggling to reproduce claims when, perhaps, they should not be bothering. Replication, for all its conceptual importance, is a messy, social business; it may be that it needs a messy, social medium.

Jim Giles is a reporter for Nature in London. He has just won an award from the Association of British Science Writers for his feature 'The dustiest place on Earth'.

1. Miquel, C. et al. *Nature* **418**, 59–62 (2002).
2. Glöckner, G. et al. *Nature* **418**, 79–85 (2002).
3. Kim, J.-H. et al. *Nature* **418**, 50–56 (2002).
4. Sriraman, S., Agarwal, S., Aydil, E. S. & Maroudas, D. *Nature* **418**, 62–65 (2002).
5. Sun, Z.-W. & Allis, C. D. *Nature* **418**, 104–108 (2002).
6. Clack, J. A. *Nature* **418**, 72–76 (2002).
7. Brittain, S. D. & Rettig, T. W. *Nature* **418**, 57–59 (2002).
8. Oka, T. *Nature* **418**, 31–33 (2002).
9. Goto, M. et al. *Astrophys. J.* **629**, 865–872 (2005).
10. Jiang, Y. et al. *Nature* **418**, 41–49 (2002).
11. Muguruma, Y. et al. *Exp. Hematol.* **31**, 1323–1330 (2003).
12. Collins, H. *Changing Order* (Sage, London, 1985).
13. Ichikawa, K., Fukugita, M. & Kawasaki, M. *Phys. Rev. D* **71**, 043001 (2005).

ON BORDER PATROL

The United States has embarked on a huge effort to try to track the H5N1 avian flu virus in birds migrating into the country. But is surveillance more urgently needed elsewhere? **Erika Check** reports.

E CHECK

On a Sunday afternoon in June, graduate student Brad Comstock is standing on the flat, soggy tundra at the edge of the Bering Sea. He surveys the network of shallow ponds stretching out in front of him, and the vast delta reaching to the horizon beyond. Nestled down along the shores of the ponds are dozens of large, migratory sea ducks called eiders. Each bird is settled on a down-lined nest holding eggs that are just days away from hatching. The air is filled with the sound of eiders honking to each other over the constant, whistling wind. "This is eider heaven," Comstock says.

But he is about to stir up trouble in paradise. Carrying a fishing net and a backpack full of test-tubes and cotton swabs, he creeps towards one of the nests. At the last moment, he hurls the net towards a mother eider. She jumps to evade him, but too late; her wings flap angrily as the net ensnares her.

Comstock rushes in, extracts the offended bird and turns her upside down. He runs a cotton swab across her cloaca to take a faeces sample, then drops the swab in a test-tube. The sample will be held briefly at a nearby camp, then shipped to the National Wildlife Health Center in Madison, Wisconsin, where technicians will test it for the H5N1 avian influenza virus.

The 7.7 million hectares of the Yukon Delta National Wildlife Refuge in Alaska may seem like heaven to eiders. But to the US government, it is the frontline in the battle to protect the country from the deadly bird flu virus.

Comstock is part of a massive effort to track the possible entry of H5N1 into the United States. Since 2003, the virus, a more lethal strain than the flu viruses that normally infect birds, has rampaged through the rest of the world. More than 200 million poultry have died of H5N1 or have been culled to prevent its spread since 2004; 132 people have died after catching the virus from close relatives or directly from birds. So far, the virus hasn't learned how to jump efficiently from person to person. But public-health officials fear that if it does, a pandemic could follow, killing millions of people.

The US government is worried about wild birds because the country is linked directly to Asia — where H5N1 first appeared — through two overlapping migratory bird flyways (see map). Every year, birds such as pintails, eiders, ducks, godwits and geese cross from Asia over the Bering Sea into Alaska. These birds mingle with other migrating groups at breeding and wintering grounds in Russia and western Alaska. No one knows which of them might be



Brad Comstock is helping the US mission to test 15,000 wild birds for the flu virus in Alaska alone.

A ROUTE FOR FLU?

As the place where two major bird flyways overlap, Alaska is under watch as a possible entry point for H5N1 into the United States.

carrying H5N1, or whether the virus might leap from such a carrier into people or into the \$29-billion poultry industry of the United States.

This year, the US departments of agriculture and the interior will lead a \$29 million effort to test wild birds for H5N1 and other avian flu viruses in Alaska and other parts of the United States. "We're at the nexus of bird migration from the Canadian Arctic to Russia and southeast Asia," says Robert Leedy, chief of the US Fish and Wildlife Service's office of migratory bird management in Anchorage, Alaska. "If H5N1 is going to be transferred in wild birds, the most likely avenue is Alaska."

Crossing continents

But some scientists have reservations about the testing programme. Many flu experts think poultry smuggling or imports, rather than migrating birds, are far more likely to bring in the virus. And others point out that it's still not clear how or whether wild birds contribute to H5N1 outbreaks in domestic poultry. Given these uncertainties, some question the decision to spend millions of dollars hunting for flu in Alaska, when the H5N1 virus is already racing across the rest of the globe. "More information is always better, so you can't complain about that," says William Karesh, director of the Wildlife Conservation Society's Field Veterinary Program based in New York City. "But it's much more important to go where the disease is in the developing countries, to see how this thing is spreading."

Until last year, no one thought that migratory birds played any serious role in the spread of H5N1. But in July 2005, a team of virologists reported that some 6,000 migratory birds had died of an H5N1 outbreak at the Qinghai Lake nature reserve in China¹. Many of the dead birds were bar-headed geese, which fly from China to India and Myanmar every year. Since that report, the H5N1 strain has been found in dead migratory birds in Asia, Russia, Europe, Africa and the Middle East. Four people even died from bird flu after collecting feathers from infected wild swans in Azerbaijan². So what role do migratory birds play in spreading H5N1 around the world?

Genetic studies may help to answer this question. This May, Ian Brown of the United Kingdom's Veterinary Laboratories Agency in

Weybridge revealed that H5N1 viruses taken from dead wild birds in Europe are very similar to H5N1 viruses found in Mongolia, Siberia and Qinghai Lake. Scientists have also reported that healthy birds in China were carrying the H5N1 strain just before their autumn migration last year³. That suggests the birds could have caused the outbreak of the virus in Europe last autumn, by carrying it to the continent from east Asia.

Other studies have implied that wild birds shuttled the virus between Europe and Africa, where H5N1 first showed up in February. A team of researchers recently suggested that

"It's much more important to go where the disease is in the developing countries, to see how the thing is spreading."

— William Karesh

migrating birds may have transmitted H5N1 to Nigeria, the first African country to report the virus⁴. The scientists sequenced genes from bird flu viruses found in chickens on poultry farms. They discovered that many of the viruses, which seemed to cluster into three genetic groups, were similar to those found on other continents, including one strain that has been found only in wild birds in Europe. What's more, the virus outbreaks in poultry were found along major bird migration corridors.

But none of these studies can conclusively show that migratory birds transmit the virus. In all cases, the wild birds themselves could have caught H5N1 from poultry or from some 'bridge' group, such as crows, jays or grackles. Officials agree that answering questions about the role of wild birds will require a lot more field work in live, migrating birds.

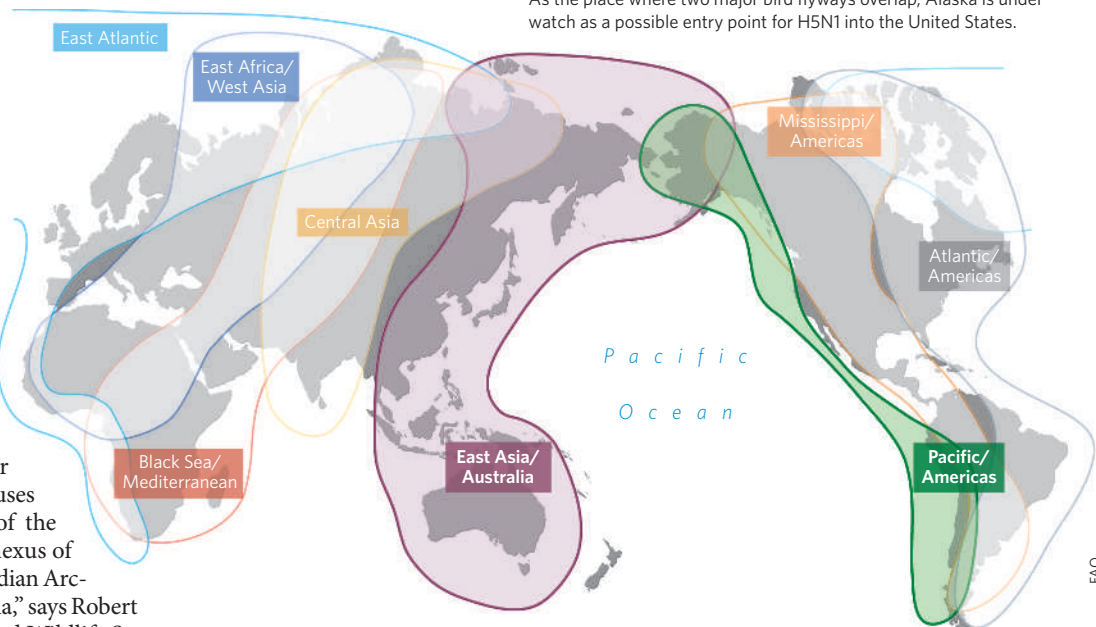
And that's why James Sedinger and his team of young biologists, including Comstock, are spending their summer swabbing birds' rears on the Yukon Delta.

Sedinger, a wildlife biologist at the University of Nevada at Reno, has been studying migratory birds in Alaska since 1977. His field camp sits at the edge of the tidal Tutakoke River, near where it runs into the Bering Sea. The camp rattles with the noise of birds at all hours in summer, when the sun sets for three hours every night, and if you don't watch your step you're likely to stumble into a mother bird sitting on a nest tucked into the grass.

This is breeding central, not just for migrating eiders, but also for black brants — chunky sea geese that fly from Alaska to points south for the winter. Over decades of work, Sedinger has snapped identifying bands on thousands of brants at Tutakoke. Some of the birds have been found to spend the winter in China, Japan and Korea. The US government's plan to head off bird flu focuses on about 29 such species that spend time in Alaska and travel to Asia. About 15,000 birds will be tested in the state. Most will be trapped live by biologists. But the Department of Interior will also test birds shot by sport and subsistence hunters and — if and when they happen — dead ones found among mass bird die-offs.

Out for the count

So far, 3,772 samples from Alaska have been tested, and none has turned up positive. But that's not surprising. Most of the samples taken from live wild birds around the world are clear of H5N1, says Ward Hagemeijer of Wetlands International, a group that coordinates volunteer surveys of migratory birds. Last winter, Wetlands International tested almost 6,000 wild birds for H5N1 along migration paths in Africa, Europe and Asia. And scientists



funded by the European Union have tested a total of 45,000 wild birds in Europe since last autumn. So far, none has turned up positive for H5N1, even though dead wild birds in all of these places have been found to be carrying the virus. Migrating birds are known to carry avian influenza strains other than H5N1. But, says Hagemeijer, "Finding H5N1 in healthy wild birds is amazingly difficult."

That perplexes wildlife-health specialists. Many of the dead or dying migratory birds found to be carrying H5N1 have been found near poultry farms, so it's not clear whether the wild birds infected the poultry or vice versa. And nobody knows how long the birds survive with H5N1, whether they are able to transmit the virus to other species and, if so, for how long they are contagious.

Hagemeijer cautions that even the most incriminating data pointing to the role of wild birds in transmitting H5N1 — the genetic study of H5N1 in Nigeria — leave some gaps. Nigeria doesn't conduct rigorous safety checks on imported poultry. So it's possible that shipments of infected birds seeded the outbreak, a point the original research team concedes. The finding that one of the Nigerian strains matches a strain found only in wild European birds also isn't convincing evidence that migratory birds were the cause of the outbreak, says Hagemeijer. This is because many more H5N1 strains have been studied from wild birds than from poultry. "If you're looking at relations between strains, you're far more likely to find a close relative in the wild bird database than in the database for poultry, because more of them have been sequenced," he says.

Needle in a haystack

Hagemeijer adds that wild birds "probably play a role, but we still haven't found the smoking gun". Other animal-health officials agree. The real question is whether wild birds can serve as a permanent reservoir for the virus, rather than simply transport it from place to place.

That is a worrying possibility because whenever the virus has a chance to mix in large groups of animals, as in the huge poultry farms of Asia, it might mutate so that it becomes lethal to people. "There are slight differences between the strains we find in wild birds, which suggests something is happening," says Scott Newman, a veterinary health specialist with the Wildlife Conservation Society who works at the headquarters of the United Nations Food and Agriculture Organization, in Rome. "But it's hard to tell whether it's happening in wild birds or in poultry."

Given all these unknowns, some are sceptical about the United States investing so much money hunting for H5N1 in live birds. Newman points out that H5N1 has swept from Asia towards Europe, and not the other way around. Perhaps, he suggests, officials should be more concerned about migration from Europe, through Greenland to the US east coast, where limited testing is ongoing. Even the biologists



Biologists are taking faecal swabs off eggs and birds' rears in the Yukon Delta in Alaska to test for H5N1.

doing the work say it's unlikely that they will intercept the virus. "We're looking for a needle in a haystack," Sedinger says.

Others are frustrated that the United States is spending so much on testing birds within its own borders while the disease continues to spread elsewhere. "We should be doing more overseas," Karesh says. "I'd like to see the United States do more outreach."

The Food and Agriculture Organization is already working with groups such as Wetlands International, the Pasteur Institute, the French Agricultural Research Centre for International Development and the Centers for Disease Control and Prevention to test birds worldwide. This January, the United States pledged to spend \$334 million in international aid for

"Wild birds probably play a role, but we still haven't found the smoking gun." — Ward Hagemeijer

countries battling influenza. And last month, the US Agency for International Development agreed to provide \$5 million to an effort called the Global Avian Influenza Network for Surveillance. Headed by the Wildlife Conservation Society, this will test birds for H5N1 in remote and poor places, such as Mongolia, that don't have their own surveillance system in place. This is the only good way to get a better handle on the dynamics of the virus in wild birds, Karesh says.

Meanwhile, biologists who normally struggle for grants to study the basic biology of migrating birds are frustrated by the huge

amounts of cash flowing for bird-flu studies. The field camp at Tutakoke, for instance, wouldn't even have been funded this summer without the bird-flu surveillance money.

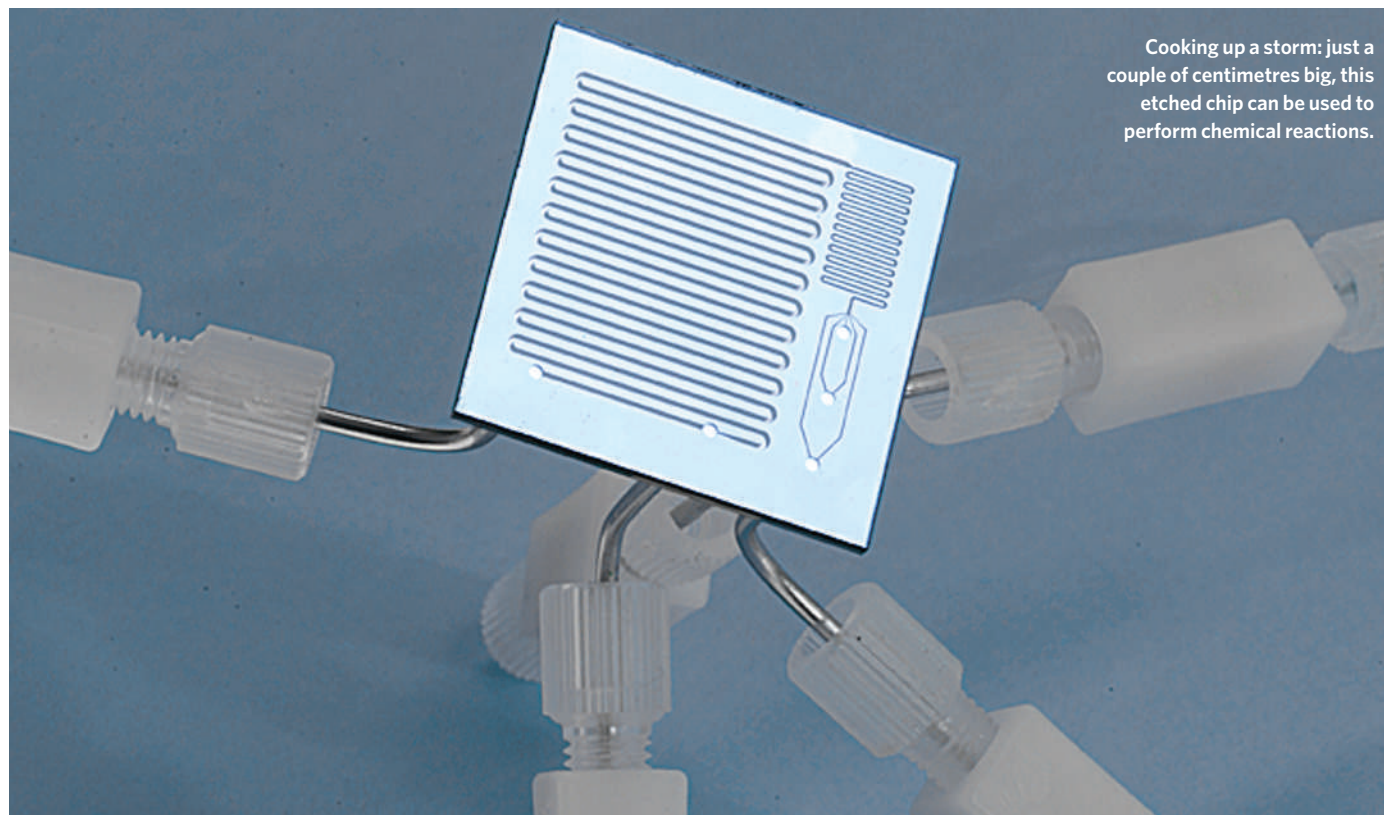
Sedinger and other bird biologists say the paucity of long-term studies on migratory birds makes it difficult to understand the spread of H5N1. It wasn't until 1995, for example, that scientists discovered that spectacled eiders spend their winters floating in holes in pack ice in the Bering Sea⁵.

So while biologists are thankful for the bird-flu surveillance money, they also wonder whether officials will take the logical next step and invest more in monitoring studies as well.

Still, biologists are happy for any chance to learn more about migratory birds — especially if it means spending long days chasing after ducks and geese on wind-blown tundra. And even if they don't track down the virus in Alaska, Sedinger, Comstock and the rest of the Tutakoke team will learn more about how viruses circulate in and between species — an important area of research given that at least two human flu pandemics in the past century began as bird viruses. "We're going to get our hands on some birds we wouldn't normally get to study," says Leedy. "That's going to lead to some good work."

Erika Check covers the biomedical sciences for Nature.

1. Chen, H. *et al.* *Nature* **436**, 191–192 (2005).
2. Eurosurveillance www.eurosurveillance.org/em/v11n05/1105-222.asp (2006).
3. Chen, H. *et al.* *Proc. Natl Acad. Sci. USA* **103**, 2845–2850 (2006).
4. Ducatez, M. F. *et al.* *Nature* **442**, 37 (2006).
5. Petersen, M. R., Larned, W. W. & Douglas, D. C. *Auk* **116**, 1009–1020 (1999).



Cooking up a storm: just a couple of centimetres big, this etched chip can be used to perform chemical reactions.

F. FRANKEL/REF. 1

A little goes a long way

Faster, safer and easier to control — chemical reactions in microreactors are taking off in the lab. Now industry is being seduced by the charms of the lab on a chip. **Jenny Hogan** investigates.

A few years ago, a productive PhD student in Peter Seeberger's chemistry lab would run three or four experiments a day. Each would be a painstaking step towards optimized conditions for a new reaction — be it making a peptide or producing a sugar molecule for use in a possible vaccine.

Since then, Seeberger's expectations have soared. Now his students have to work ten times as hard. The 120 reactions that formed the basis for one recent publication¹ were completed in three afternoons.

It's not that Seeberger, at the Swiss Federal Institute of Technology in Zurich, has become a slave-driver. Rather, he has updated his lab equipment. He is working with a collection of microreactors — each one a miniature lab on a chip. The reagents are stirred up again, and again, in channels less than a millimetre in diameter, until the students get the results they need.

"People in my lab are very excited about this. It gives you time to do more chemistry," says Seeberger. "I always say that microreactors will be chemists' round-bottomed flasks for the twenty-first century."

The humble glass flask is a chemistry icon, used since alchemists tried to turn base metals

into gold. But microreactors promise to make chemistry faster, cleaner and yield purer products. They might also open the door to syntheses not previously feasible on a large scale, and make dangerous — even explosive — reactions safer.

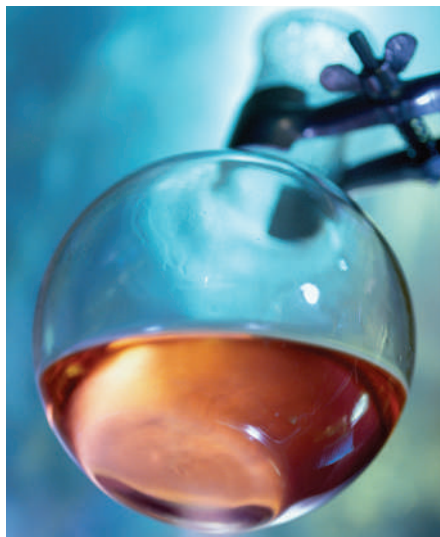
The technology has grown over the past two

decades — a convergence of the miniaturization of chemical and biological analysis techniques and the engineering of computer chips. Seeberger's chips (pictured above) are typical of what is possible. Just a couple of centimetres big, they feature tiny channels etched into silicon. Chemicals are injected into the device and they react where they merge. The bends in the channels help force the reagents to mix, and the length of the channels and the flow rate determine the reaction time. With reaction volumes measuring just microlitres, conditions such as pressure and temperature can be precisely controlled and quickly changed.

"You only have to run the system until you have one drop of product coming out of the end. You'd spend the longest time walking downstairs to the spectrometer to analyse it," says Graham Sandford, a chemist at Durham University, UK, who has used microreactors in his lab².

Now the technology is also making the leap into industry. Microreactors performing cleaner and safer reactions could push the batch vessels used in the synthesis of some compounds — including drugs — into retirement. Ultimately, the devices could end up integrated into drug discovery³.

"For me, the important message is that the



End of an era: could lab-on-a-chip technology spell the end for the round-bottomed flask?

technology can be applied on an industrial scale," says Volker Hessel, vice-director for research and development at the Institute of Microtechnology Mainz (IMM), which has collaborated with hundreds of chemical companies on microreactor projects.

In industry, high-value products that are typically produced in small batches, such as pharmaceuticals and fine chemicals, have much to gain from the extra flexibility offered by microreactors. Scaling up a lab procedure to batch production can sometimes require redesigning a chemical reaction from scratch. "In flow world, you just run a reaction longer," says Tony Wood, head of discovery chemistry for Pfizer in Sandwich, UK.

Wood says that Pfizer is only just beginning to explore the possibilities that the technology offers, but he hopes microreactors will change the rules for his chemists. Reactions that they now have to avoid because they are difficult to run in large volumes might become accessible. "What's interesting to me is the opportunity to pursue fields such as electrochemistry or photochemistry," says Wood. "That would enable us to functionalize molecules in a quite different way from mainstream transformations."

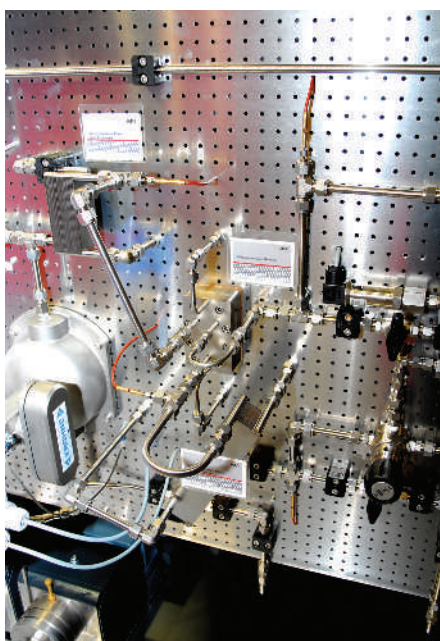
Compound interest

But there are still some engineering troubles to be overcome. Precipitates are a problem in any reaction process, but in the tiny channels and chambers of a microreactor, clogging is an ever-present danger. "Solids are problematic and if you can avoid them you will try," says Hessel. Researchers in the field cite the growing literature on systems that can handle solids as evidence that the problem of clogging will, with time, be conquered. The challenges of performing reactions in which different steps in a synthesis require different solvents are also being dealt with, say industry insiders.

As a result, industrial interest in microreactors is spreading fast, says Hessel. "It's nearly all the big names in chemistry," he notes. One idea that industry has been quick to latch on to is safety. For example, reactions that release a lot of energy may be controlled in a small flask, but risk exploding in a larger batch vessel where excess heat is harder to dissipate. With a microreactor, scaling up the reaction safely simply requires running more devices in parallel.

This encouraged Xi'an Huian Chemical in China to approach Hessel's team at the IMM to set up a microreactor plant to synthesize nitroglycerine. It took only five months for the team to get the plant running and, since September 2005, it has been producing nitroglycerine for use as a treatment for heart disease.

Switching from batch to flow chemistry is not just about refitting a plant or lab, it often requires a change of mindset. "You are in a



Scaled down: this fine-chemical process plant features micro mixers and heat exchangers.

small, innovative team in an established company that has more than 100 years' experience in chemical production and you want to change things — there are some barriers beyond the technical," says Dominique Roberge, head of a project to evaluate microreactors at Lonza, a Swiss company that manufactures intermediates for the drug industry.

Even Seeberger was not an immediate convert. After learning about microreactors in 2001, he decided to investigate further in his lab at the Massachusetts Institute of Technology (MIT), where he was then based. He was fortunate that Klavs Jensen in MIT's chemical-engineering department was already making microreactors. But Seeberger wasn't impressed at first. "When Klavs showed me his devices, they looked like toys. I thought they would be useless, that you wouldn't be able to make enough material." But when he came to try it, he found that running a microreactor for a day produced 100 grams of material — far more than his students are ever likely to need for biological tests.

Steven Ley, a chemist at the University of Cambridge, UK, is another who threw out many of his round-bottomed flasks after building a flow-chemistry lab. He welcomes the opportunity to do chemistry differently. For example, he says, some flask reactions have to be carried out at -195°C , the temperature of liquid nitrogen, to prevent 'overcooking' the reactants, but they can be performed at room temperature in microreactors. This would make them economical for industry.

Even for existing batch reactions, microreactors can offer an attractive alternative.

"The question of whether microreactors are going to be used in the future, I think this is already answered 'yes'."

Dominique Roberge

The flow inside their small channels is better behaved than in a large vessel, and so is more reproducible. In the drug industry, if a batch doesn't match the specifications approved by regulators, it has to be thrown away. "Many batches are lost and very often it's the culmination of several months' work," says Brian Warrington, former vice-president of technology development for GlaxoSmithKline, UK. "It's a big commercial problem."

That was one reason why Warrington pushed the idea of microreactors at GlaxoSmithKline in the late 1990s. The other was the idea of 'closing the loop'. Warrington says the ultimate goal is to have a microreactor pumping its product straight into a cell-based assay, which is hooked up to provide feedback to a computer controlling the synthesis of the next product to be tested.

Recipe for success

In his Cambridge lab, Ley is experiencing the drug industry's growing interest in flow chemistry directly. Companies are clamouring to send people to the lab for training, he says. One of his PhD students collaborates with Chris Selway, one of the drug-discovery chemists at Pfizer charged with evaluating the technology's potential for the firm.

So far, Selway remains cautious about microreactors. "We are seeing lots of claims in the literature about how good flow chemistry is and, as a company, we want to be involved in that," says Selway, "but by no means is it a tool that's going to radically take over from batch chemistry."

Industry still has to work out the economics of microreactors. Lonza began operating a pilot plant using microreactors in March and Roberge has analysed 83 reactions performed by the company⁴. He found that half could benefit from being carried out in microreactors, although solids in many of these reactions reduced that number to 16. His preliminary economic analysis suggests that the cost of building and commissioning the microreactor plant will be comparable to a batch system of similar throughput — around €250,000 (US\$316,000). The main hope for future cost savings, he says, is if microreactors can deliver improved yields and so use lower amounts of raw material.

"The question of whether microreactors are going to be used in the future, I think this is already answered 'yes'," says Roberge. "The open question is what per cent of the market in fine chemicals they will take."

Jenny Hogan is a reporter based in Nature's London office.

1. Ratner, D. M. *et al. Chem. Commun.* 578–580 (2005).
2. Chambers, R. D. *et al. Lab Chip* **5**, 191–198 (2005).
3. Dittich, P. S. & Manz, A. *Nature Rev. Drug Discov.* **5**, 210–218 (2006).
4. Roberge, D. M., Ducry, L., Bieler, N., Cretton, P. & Zimmermann, B. *Chem. Eng. Technol.* **28**, 318–323 (2005).

For more on lab-on-a-chip technology see the Insight on pages 367–418 of this issue.

Hungary: academy is not obsolete or discriminatory

SIR — We were puzzled by your News story “Hungary’s science academy slammed as ‘obsolete’” (*Nature* **441**, 1034–1035; 2006), because we had previously provided you with information that contradicts it. Contrary to your story and the statements made by people cited in it, our academy does not discriminate against scientific achievement obtained outside Hungary.

To cite the relevant part of our byelaws when conferring a Doctor of Science (Doctor of the Academy) title, from the www.mta.hu website: “When evaluating results arising from research done in an advanced industrial country, articles can be included fully in the impact factor total only if the applicant is the first or last mentioned co-author. If not, articles can be included in the impact factor or citation index total only to the effect of 25%. If the applicant competes for a Doctor of the Academy title on the basis of research done in a foreign country scientifically as advanced as Hungary, or less advanced than Hungary, reaching the domestic minimal measure is the minimal condition of allowing him/her to compete for the title.”

By quoting people who call the Hungarian Academy of Sciences “obsolete” you are unjustly offending almost two-thirds of its members, and indeed three-quarters of our Doctors of the Academy, elected after 1990, not to mention more than half the academy’s research staff, who started their scientific careers after that date.

We are troubled that you ignored the information we provided to your reporter before the article was published, and instead featured overwhelmingly the views of people who had previously denounced the academy and the scientific community in Hungary, and had met with rebuttals. We find your News story tendentious and partial, at odds with *Nature*’s widely respected objectivity and editorial integrity.

György Fábrí

Communications and Science Policy,
The Hungarian Academy of Sciences,
Roosevelt sq. 9, H-1051 Budapest, Hungary

Hungary: academy needs more than internal reform

SIR — Your News story “Hungary’s science academy slammed as ‘obsolete’” (*Nature* **441**, 1034–1035; 2006) describing some of the practices of the Hungarian Academy of Sciences, with respect to the requirements for its doctoral degree, only touches the tip of the iceberg. There are many chronic problems with the academy (see *Nature* **427**, 94–95, 2004), such as the ageing leadership, lifelong

tenure and funding allocations heavily favouring academy members and their institutions.

Although I did, as your News story reported, return to Hungary after 10 years of working in the United Kingdom and the United States, I stayed for only one year because of these outdated practices. I called for an internationally competitive science-management system and spoke out against the state-funded privileges and the lifelong compensation system the members of the academy enjoy. Two days later, I was told by the head of my department at Semmelweis University to look for another job.

A radical change in Hungarian science administration is overdue. The recent removal of the criticized sections from the academy’s webpage is not going to change anything. I seriously doubt that the new ‘internal reform committee’ of the Hungarian Academy of Sciences is what the country needs.

Csaba Szabo

Department of Surgery, University of Medicine
and Dentistry of New Jersey, Newark,
New Jersey 07103, USA

Quest for seed immortality is mission impossible

SIR — In your News story “Doomsday food store takes pole position” (*Nature* **441**, 912–913; 2006) you report the establishment of a “doomsday seed bank from which the genetic riches of Earth’s food crops could, if necessary, be reconstituted”. Unfortunately, this statement is wishful thinking.

As managers of one of the largest *ex situ* gene banks in the world, we have been surprised during the past few weeks by the increased number of comments by visitors about the possibility of seeds being kept viable for millennia if stored at sufficiently low temperatures. There are very few reported cases in which seeds were able to germinate after many decades or even more than a century of storage: most cannot survive long-term storage.

One prominent example is provided by a herbarium sample of the legume *Albizia julibrissin*, which had been collected in China in 1793 and was deposited in the British Museum. The material was accidentally wetted in 1940 during firefighting, resulting in the germination of several seeds at least 147 years old (*Nature* **149**, 658; 1942). There are a few other documented examples where seeds have been able to germinate after 100 years (details available on request from the authors), but these are rare exceptions. The suggestion that seeds can survive after thousands of years is wrong, and detrimental to conservation funding efforts.

Undoubtedly, it is important to keep safety

duplicates for the estimated 6 million seed samples that are kept in *ex situ* gene banks worldwide. To this end, the food store featured in your News story is an important contribution. But the vast majority of plant seeds cannot be stored for more than 40 years without losing germination vigour, however low the temperature and however thick the concrete walls of the vault. Hence, gene banks have no other choice but to preserve their material by growing plants in greenhouses, field plots or in laboratories at regular intervals. This process requires manpower, knowledge and rigorous quality management. It is also the bottleneck in almost all the gene banks so far established: the seeds they contain urgently require regeneration, otherwise they will die.

Andreas Graner, Andreas Börner

Federal *ex situ* Genebank, Leibniz Institute
of Plant Genetics and Crop Plant Research,
06466 Gatersleben, Germany

No place for secrets in scientific research

SIR — I heartily endorse the call in your Editorial “Illuminating the black box” (*Nature* **442**, 1; 2006) for transparency in describing materials and reagents used in scientific papers. As editor-in-chief of the *Journal of Comparative Neurology*, one of the largest and oldest neuroscience journals, I have for some time required authors to provide full information on the origin and specificity of all reagents, especially antibodies (C. B. Saper and P. E. Sawchenko *J. Comp. Neurol.* **465**, 161–163; 2003).

In addition to the criteria raised in your Editorial, for “specificity and utility” of an antibody, the chemical structure against which the antibody was raised must be known for an experiment to be replicable. Commercial availability of a reagent does not satisfy this need, as the manufacturer may change the nature of a reagent, or go out of business. Many commercial firms have taken to declaring the chemical structure against which the antibody was raised as ‘proprietary’, in other words, a trade secret. I tell authors that ‘proprietary’ is another term for ‘not fit for serious scientific work’, and our journal will not publish work that is done with such reagents.

I applaud *Nature* for insisting on full disclosure about reagents. It will be hard work to implement these policies, as many reviewers are not yet sufficiently aware of or concerned by this issue, but the integrity of science depends on it.

Clifford B. Saper

Department of Neurology and Program in
Neuroscience, Harvard Medical School,
and Beth Israel Deaconess Medical Center,
Boston, Massachusetts 02215, USA

BOOKS & ARTS

Design flaws

Destroying the argument that intelligent design has a scientific basis.

Intelligent Thought: Science Versus the Intelligent Design Movement

edited by John Brockman
Vintage Books: 2006. 272 pp. \$14

John Tyler Bonner

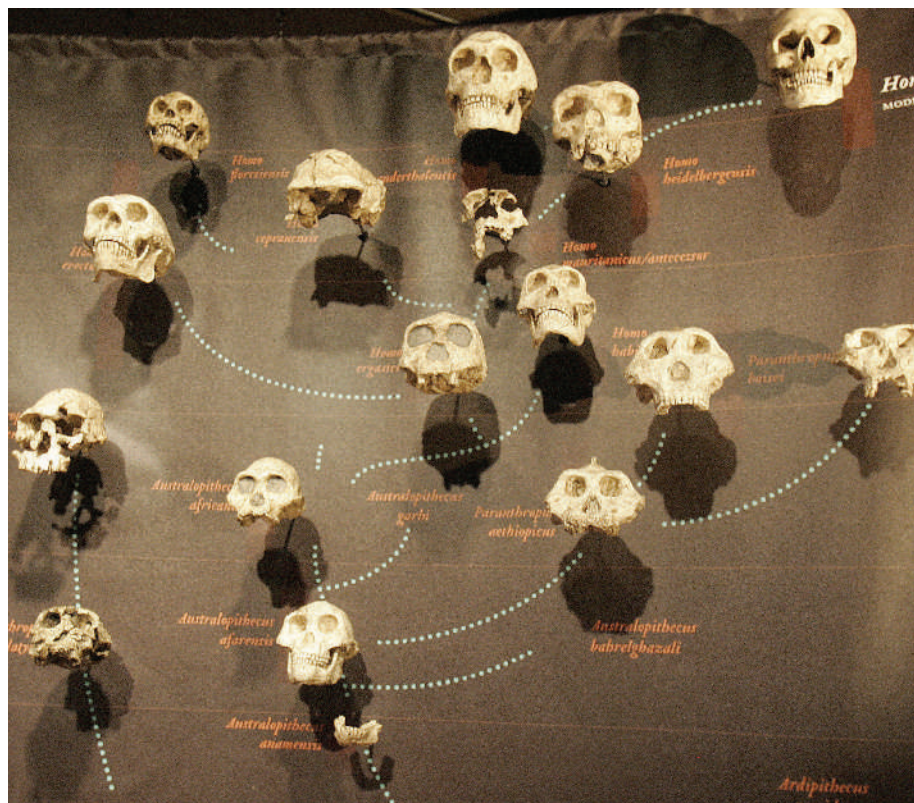
'Intelligent design' — the idea that life was shaped by an intelligent creator — has caused a stir in education circles in the United States. Its protagonists have pushed it as an alternative explanation to evolution and are demanding equal time for it in the classroom (see *Nature* **416**, 250; 2002). The idea is considered ridiculous by almost the entire scientific community, and so far the legal challenges to introduce it in schools have failed.

Advocates of intelligent design have tried to invent a new definition of science that includes religious beliefs, with the clear intent of circumventing the strictures separating church from state, and allowing creationism to be taught in schools. This is an important constitutional and legal issue, and a serious question for the US educational system, but it is hardly a scientific one.

John Brockman's edited volume *Intelligent Thought* is largely a series of essays by scientists that make clear, often eloquently, how untenable the scientific basis of intelligent design really is. In some ways it is like shooting fish in a barrel. When giving his decision at the trial in which the school board in Dover, Pennsylvania, was challenged for including intelligent design in the biology curriculum (see *Nature* **439**, 6–7; 2006), federal judge John Jones described the argument that intelligent design is science as a "breathtaking inanity". His ruling is reprinted in part at the end of the book.

If intelligent design has anything to say in its favour, it is that it spawned this book. Many of the essays are fascinating and fun to read, and tell us something new.

In the opening essay, Jerry Coyne directly attacks intelligent design by showing how it is refuted by conventional evolutionary biology. The approach may be somewhat standard, but he does a splendid job of devastating the central arguments of intelligent design's proponents. In no way can intelligent design qualify as science, and he makes the case in the manner of a canny lawyer. Several of the other essays in the book take up many of his themes and arguments, and the end result is an over-



Use your head: anthropologists have made an overwhelming case for evolution.

whelming indictment of intelligent design. It is not science because it is not falsifiable. Many of these arguments are also to be found in Judge Jones's decision; one might call them the core arguments as to why intelligent design is not science. They are correct, and devastate the arguments for intelligent design.

The book also includes other approaches to the matter, two of which I found particularly effective and worthy of mention. One follows a specific evolutionary study in some detail, which, like Darwin's method in *The Origin of Species*, gives an outpouring of all the abundant evidence, smothering objections and disagreements. It comes from Tim White, who provides a riveting tale of what physical anthropologists have done to puzzle out the ancestry of modern humans. He takes us to the plains of Ethiopia and tells us how little fragments of bone are pieced together and how the result is interpreted. He then shows where this find fits in with the many other discoveries in Africa and other parts of the world.

So many skulls are now known that human ancestry can be traced with a degree of precision quite impossible a few years ago. He makes the case for evolution using irresistible and overwhelming detail — the literary equivalent of saturation bombing — so it is difficult to come to any other conclusion.

Frank Sulloway, in contrast, uses a historical approach to the dichotomy between creationism and evolution. He shows that Darwin was originally a creationist, and remained so during his entire voyage on the *Beagle*. As Sulloway reminds us, Darwin seriously contemplated studying for the ministry while at the University of Cambridge and showed no signs of deviating from church doctrine. Furthermore, he was greatly inspired by the theologian William Paley, who was more or less the father of intelligent design. It was only when Darwin returned from his voyage, and began to examine the specimens he had collected and consider his observations on the Galapagos and elsewhere, that he realized

the possibility of evolution and the transmutation of species. And from this idea sprang natural selection. It is illuminating and convincing to see how Darwin, in his intellectual maturing, changed from intelligent design to evolution.

Intelligent Thought is a book for scientists;

that is, for those who see evolutionary biology as a science. If you are a creationist you will be unmoved; there is no point in looking at the evidence. ■

John Tyler Bonner is in the Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, USA.

Diary of a weed

Seed to Seed: The Secret Life of Plants

by Nicholas Harberd

Bloomsbury: 2006. 320 pp. £16.99, \$24.95

Anthony Trewavas

Diaries are objects of intense fascination. Not only do they provide unique insight into the private thoughts and attitudes of the author, rarely conveyed by autobiography, but they can become important historical documents. Famous diarists such as Samuel Pepys in the seventeenth century and John Colville during the Second World War described the intimate workings of British government during periods of enormous social upheaval and war. Others, like the British politician Alan Clark, who consigned to their diary assessments of their political peers they would never voice in public, provide highly entertaining reading.

Diaries by working scientists are rare, so with the above thoughts in mind, I welcomed the chance to read Nick Harberd's contribution to the genre. Harberd is a plant molecular geneticist at the John Innes Institute in Norwich, and kept 'jottings' throughout 2004. Apart from the inevitable descriptions of seasonal changes in weather and wildlife and outlines of conversations with his two children, it slowly becomes apparent that the real subject of this diary is *Arabidopsis*, a weed with a very short life cycle (seed-to-seed) of some 4–6 weeks. These characteristics have made it the plant of choice for an enormous amount of gene identification and developmental analysis. Harberd uses the diary as a vehicle both to describe simple aspects of plant development and to structure his own research on *Arabidopsis* in 2004. The life cycle of a few wild specimens of *Arabidopsis* (in contrast to lab-grown versions) is also lovingly described — it ends when they provide a first-class meal for some roaming slugs!

Harberd records the thoughts, guesses, assessments, excitement, obsessions, publication problems and productive experiments that are the meat and drink of any experimental scientist. His significant achievement of 2004 was the identification and sequencing of the crucial *rht* (reduced height) gene. The 'green revolution' of the 1960s and 1970s resulted from the introduction of this gene into wheat and rice by Norman Borlaug. The resulting dwarf cereals had grain yields more than three times greater than their parents,

and turned Mexico, China and India into net cereal exporters. An estimated 1 billion humans were saved from the savagery of slow starvation and premature death that some scientific Jeremiahs had long predicted as their inevitable fate.

The identification of this gene thus represents the final chapter in a story that vividly illustrates the value of technology in ameliorating human problems. We are technological animals. All human activities (including technology) have costs as well as benefits, but our capacity to rise to the challenge of complex problems is frankly what makes us human and makes living worthwhile. To deny this, to

deny the value of the advance of science, as some currently do, is to deny our humanity.

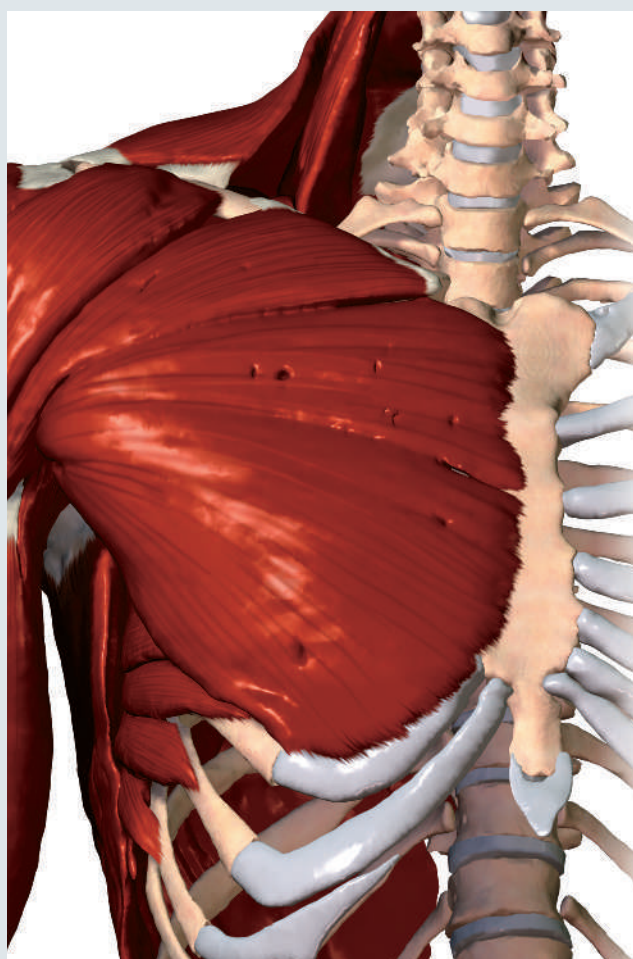
Harberd's diary makes for easy and enjoyable reading for a plant biologist like myself, but I gradually began to wonder who its wider audience might actually be. Commentaries on seasonal changes in weather and nature form the basis of many published Edwardian diaries and are often better illustrated. The introduction makes clear that the author thinks the diary is written for non-scientists, but I was left wondering who, if any, of my non-scientific friends would make sense of it all. There is enormous interest in Britain in plants of all kinds, but from a gardening and horticultural perspective only. So a diary about a weed is unlikely to gain a sympathetic readership here. Despite the author's view of his readership, I think that other non-botanical scientists are the most likely audience and that many animal scientists and microbiologists would find pleasure in learning how the other half lives and works.

So is this diary instead a charm offensive from genetics, a subject often painted as based on grim-visaged determinism? Is it the velvet glove of the intensely human diarist hiding a genetical iron fist underneath? Harberd

The body bared

From the early Renaissance on, our drive to discover, uncover and map the unknown has found rich pickings in the human body. Laying it bare from bone to skin became a mania among medics and artists alike, starting with Leonardo da Vinci and gaining momentum with the likes of Jacopo Berengario da Carpi, Andreas Vesalius, and later William Cheselden, John Hunter and today's digital modellers.

Human Anatomy: Depicting the Body from the Renaissance to Today by Benjamin A. Rifkin, Michael J. Ackerman and Judy Folkenberg (Thames & Hudson, £17.95) dissects this history with elegance — and with the aid of 320 stunning illustrations that cover the gamut of styles, from Baroque surreality to the drily detached.



ponders on the general perception of science as unemotional, requiring publications that omit feelings, wonder and fascination at the complexity of life in all its ramifications in favour of bland objectivity. But in that case, it doesn't help to read that our *raison d'être* is our genes, a view I profoundly disagree with but an error commonly voiced by geneticists. Cells, organisms, tissues and populations are

all complex integrated systems; genes are part of these systems, but what makes us is the whole, not any one part.

This is not to decry Harberd's effort. Anything that reveals to the public that under the skin scientists are human like anyone else is to be welcomed; only scientific knowledge is blandly objective, but therein lies its strength. Harberd's diary is, I hope, the harbinger of

others, and should be regarded as a valuable and successful experiment. Biographies of scientists are common, but diaries containing day-to-day thoughts and speculations are far more attractive. If only Newton, Einstein or Darwin in later life had obliged. ■

Anthony Trewavas is at the Institute of Cell and Molecular Biology, Kings Buildings, University of Edinburgh, Edinburgh EH9 3JH, UK.

Playing the numbers game

Does Measurement Measure Up? How Numbers Reveal and Conceal the Truth
by John M. Henshaw

Johns Hopkins University Press: 2006.
248 pp. \$26.95, £14.63

David Colquhoun

Now, more than ever, we are besieged by numbers, targets, league tables and non-stop auditing. We are inundated with management gobbledygook. Most of these activities produce numbers, but what do all these numbers mean? That is the question asked in John Henshaw's book *Does Measurement Measure Up*, and it is undoubtedly a timely question.

The book starts by quoting Lord Kelvin: "When you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind." What scientist could doubt that? Realism soon creeps in though: "The darker side of measurement ... manifests itself in a variety of ways, perhaps nowhere more insidiously than in 'measurements of the mind'." Henshaw proceeds to take apart the ludicrous idea that a person's worth can be measured by a single number. This is a book that is intended for a wide audience, so the discussion is not very technical. Interesting stuff, such as a good discussion of the nature-nurture problem, is missing almost entirely (perhaps because the author is an engineer not a geneticist).

Henshaw's *bête noire* is the weighted mean. Nothing wrong with it, of course, when the weights have some proper basis. The problem arises when the weights are just produced from a hat to combine separate measurements into a single number. That means that you can get almost any number you want by adjusting the weights: a classical garbage-in, garbage-out problem. Henshaw illustrates this beautifully by his evaluation of the US system used to rank universities. Why give 25% weight to peer assessment and 10% to expenditure per student?

Using other weights would produce different outcomes. His account of the man who was hired by the University of Tulsa, Oklahoma, to improve its ranking is chilling, but many universities now waste money in that way.

In some ways UK university rankings are even worse. Here the rankings involve measures of teaching quality. That sounds entirely reasonable until you look at where the numbers come from. They come from the Quality Assurance Agency (QAA), an organization that costs £11.5 million (US\$21 million) annually,



I am not a number: evaluating people by numbers is often meaningless, but we still do it.

but which seems to think that the quality of teaching can be measured numerically without bothering about what is being taught. This leads to absurd results, like university courses in homeopathy (yes, there are some) being given near-perfect scores when they should really be referred to the Office of Fair Trading for describing as science a subject that is more akin to magic. The antics of the press in their attempts to rank universities are a more-or-less honest attempt to make money. The reification imposed by the QAA is an intellectual disgrace. As so often, it is a case of '*Quis custodiet ipsos custodes?*' Or possibly, a case of, 'those who can, do; those who can't, assess'.

Now back now to the book. The best bits are the discussions of IQ and of rankings. They deal, if incompletely, with the absurdities, and

the real social harm, that can come from trying to measure what is not measurable (at least by a single number). Other chapters discuss ranking in sports, and the sorts of measurement that can be made in business management, magnetic resonance imaging, the genome and global warming. These are subjects of great interest to many people, but I found their treatment disappointing. The chapter about ranking in sports will not mean much to anyone who does not avidly follow baseball results, and to follow one of the author's examples of climate measurement you have to be familiar with summer in Las Vegas. More seriously, though, there is no critical analysis of whether the measures used in business management, genetics and climate studies actually measure what they purport to. I would really like to know how much management-speak is just pretentious gobbledygook, but this book didn't tell me. Systems biology, an area in which clever computations are often based on dubious, or even non-existent, numbers, deserved a mention but doesn't get it.

The style of writing is easy, but it lacks passion. IQ advocates are gently taken to task, and journalists mildly castigated. I suspect that Henshaw lets off too easily the main villains of the piece, namely academics themselves. They (some of them, anyway) provide the box-ticking and paper-collecting for agencies such as the QAA. It is not unknown for some of us to exaggerate the significance of our findings, partly, of course, because of fear of the next Research Assessment (yet another case of attaching dubious numbers to complex activities). And as they get older, too many academics seem happy to take refuge behind a façade of management-speak in the style so beautifully described by Michael O'Donnell in his *A Sceptic's Medical Dictionary* (BMJ Books, 1997) as "decorated municipal gothic".

Numbers, as Henshaw's book says, are often abused, but it is academics who must shoulder much of the blame for that. ■

David Colquhoun is in the Department of Pharmacology, Gower Street, University College, London WC1E 6BT, UK.

IMAGE 100/ALAMY

SEMICONDUCTOR PHYSICS

Magnetic manipulations

Nitin Samarth

A deft technique allows magnetic atoms to be placed one by one in a semiconductor crystal. It's a further step towards an ambitious goal: a computer chip that might simultaneously store and manipulate data.

On page 436 of this issue, Yazdani and colleagues (Kitchen *et al.*)¹ show how a scanning tunnelling microscope can be used to assemble, atom by atom, a magnetic semiconductor. The result is an atomic-scale laboratory that allows an unprecedentedly clear picture of the electronic energies and magnetic interactions in the material concerned — manganese-doped gallium arsenide (refs 2, 3 and references therein). Crucially, the authors confirm predictions^{4,5} that the ferromagnetic interaction between pairs of magnetic atoms varies with their crystallographic orientation (their relative positions in the semiconductor lattice). That signals new ways to engineer, and possibly to enhance, the magnetic behaviour of ferromagnetic semiconductors — potentially with significant consequences for the burgeoning field of semiconductor spintronics.

Semiconductors and ferromagnets (the latter are materials, such as iron, that become permanently magnetized when placed in a magnetic field) form the backbone for much of contemporary information technology. These two classes of material carry out very different functions: semiconductor microprocessors are used for logic circuits (the actual 'computing' part), whereas ferromagnets are used for storing and accessing information on hard disks. Together, they allow us to process quantities of data that would have seemed preposterous even a few decades ago.

The alternative, fledgling approach to information technology known as semiconductor spintronics⁶ aims to weld the disparate advantages of semiconductor physics and magnetism into a unified framework. The ultimate goal is to create new breeds of device, such as chips that integrate logic and storage. Magnetic semiconductors like those investigated by Kitchen *et al.*¹, which incorporate magnetic atoms in a relevant semiconductor, could well become the building-blocks for this technology.

Interest in manganese-doped gallium arsenide — $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ — to use the standard technical nomenclature — was sparked by the discovery² in 1996 that this material becomes ferromagnetic at modest concentrations of manganese atoms (x greater than about 0.015).

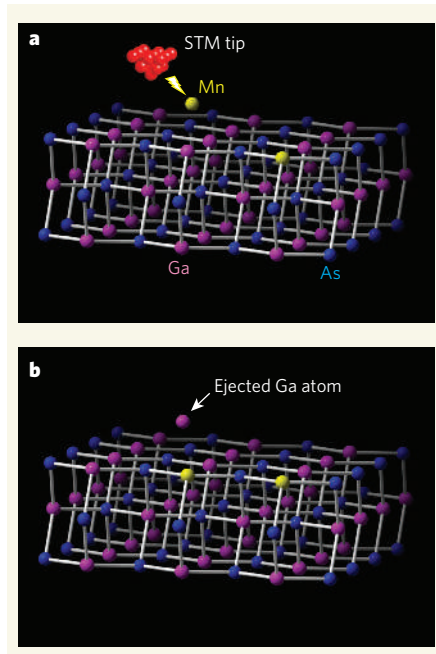


Figure 1 | How to assemble a magnetic semiconductor. **a**, In Yazdani and colleagues' experimental set-up¹, an STM tip is positioned close to a weakly adsorbed manganese (Mn) atom on the surface of a manganese-doped gallium arsenide (GaAs) crystal. Applying a voltage pulse pushes the Mn atom onto a Ga lattice site, ejecting the Ga atom onto the surface. **b**, The result is a pair of bonded manganese atoms whose spins align, introducing ferromagnetic properties in the GaAs semiconductor.

Gallium arsenide has immense technological importance in consumer electronics, being used, for example, in lasers for DVD players and high-frequency electronics for mobile phones. So the discovery of magnetic properties tantalized researchers with the prospect of integrating magnetism with a mature semiconductor technology.

Unfortunately, technological applications using $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ are still a remote prospect. Ferromagnetism comes at a heavy cost: incorporating manganese atoms at the requisite concentrations creates many defects in the crystal structure that severely degrade the

semiconductor's electronic properties. In such semiconductors, current flow is caused by the movement of 'holes' that mark the absence of a valence electron, equivalent to the motion of a positive charge. In ferromagnetic $\text{Ga}_{1-x}\text{Mn}_x\text{As}$, the mean free path for a hole is usually only a few ångströms.

Nonetheless, the presence of ferromagnetism — which arises when the atomic spins in a material line up in the same direction — at relatively low manganese concentrations raises fundamental questions for condensed-matter physics. Experiments show that manganese atoms act as 'acceptors', taking up electrons and so contributing holes to the crystal structure. The Curie temperature (the temperature above which atomic spins de-align and a material ceases to be ferromagnetic) increases for $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ with the density of both manganese atoms and holes. According to a popular view, the localized spins of manganese atoms orient the spins of the itinerant 'sea' of holes in which they are immersed³. These polarized holes in turn influence the spin orientation of manganese atoms, creating the ferromagnetic state.

But are these holes as free to move around as in a ferromagnetic metal such as iron? Or are they instead energetically constrained to move sluggishly in an impurity band? What exactly is the underlying electronic structure of $\text{Ga}_{1-x}\text{Mn}_x\text{As}$, and how does it contribute to ferromagnetism? Disorder and defects in ferromagnetic $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ often smear experimental signals, thwarting attempts to obtain unambiguous answers to these questions.

One way to clarify the microscopic nature of interactions in $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ is to study them in the magnetically 'dilute' regime, where the manganese concentration is low and defects are minimized. The first experiments to map out the local electronic states around isolated manganese atoms in $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ used scanning tunnelling spectroscopy of magnetically dilute thin-film crystals that had been cleaved to reveal a clean surface⁷. But attempts to extend these studies to interactions between pairs of manganese atoms on the cleaved surface were stymied by inherent limitations of

the experimental set-up, such as the broadening of the experimental signal caused by sub-surface manganese atoms⁸.

Kitchen *et al.*¹ take a very different approach. Rather than having an assembled crystal with a given distribution of magnetic atoms, they use the probe of a scanning tunnelling microscope (STM) to position manganese atoms at will in lattice positions on a gallium arsenide surface. The location of the manganese atoms near the surface is essential for clear experimental signatures of the interaction between neighbours.

Since the first demonstrations of STM-aided atom manipulation on surfaces⁹, the technique has matured into an exquisite tool for the experimental 'quantum mechanic', providing elegant physical pictures of the wave-like properties of electrons in artificially constructed quantum systems. An STM has recently been used to control and study fundamental magnetic interactions between complex groups of atoms on surfaces¹⁰. But useful information about magnetic interactions in ferromagnetic semiconductors such as $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ cannot be gained simply by placing magnetic atoms on the surface; they must be properly incorporated in the crystal lattice, replacing gallium atoms.

Kitchen and colleagues show that this can be done by applying voltage pulses through the STM tip at the location of a weakly adsorbed manganese atom (Fig. 1). This process mysteriously nudges the manganese atom into a site occupied by a gallium atom while ejecting the gallium atom from the lattice. The phenomenon can be exploited systematically to place pairs of manganese atoms at precise locations, so that the interaction between them can be measured as a function of their separation and their geometrical placement relative to the crystal.

Evidence for an interaction between the manganese atoms arises from the observation of energy splitting in the electronic states of the paired atoms. If the spins of two atoms are aligned ferromagnetically (that is, in the same direction), the rules of quantum mechanics allow the ground states of the individual atom systems to couple to form two states, a 'bonding' and an 'antibonding' state, of different energy. The Pauli exclusion principle tells us that each state can be filled by only one hole. Thus, the energy splitting provides a direct measure of the degree of interaction between two atoms that leads to a ferromagnetic ground state — the greater the interaction, the greater the splitting.

Remarkably, Kitchen *et al.*¹ provide clear experimental evidence for inherent ferromagnetic coupling between pairs of manganese atoms even when there is no ready supply of holes from the rest of the $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ structure. That might force us to rethink how ferromagnetism arises in this semiconductor. The authors' measurements also show that the coupling varies with the geometrical

placement of the atoms within the crystal lattice, suggesting new ways to engineer a $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ crystal possessing enhanced magnetic interactions.

The implications of this work go beyond $\text{Ga}_{1-x}\text{Mn}_x\text{As}$. One might, for example, imagine similar experiments on more challenging materials such as gallium nitride — used in the blue-light-emitting diodes central to next-generation disk technology — laced with manganese. Whatever the next step, exquisitely accurate atomic manipulation such as that achieved by Kitchen and colleagues¹ could prove crucial to resolving outstanding issues in semiconductor spintronics. ■

Nitin Samarth is in the Department of Physics,

Penn State University, 104 Davey Laboratory, University Park, Pennsylvania 16802-6300, USA. e-mail: nsamarth@psu.edu

1. Kitchen, D., Richardella, A., Tang, J.-M., Flatté, M. E. & Yazdani, A. *Nature* **442**, 436–439 (2006).
2. Ohno, H. *et al. Appl. Phys. Lett.* **69**, 363–365 (1996).
3. Macdonald, A. H., Schiffer, P. & Samarth, N. *Nature Mater.* **4**, 195–202 (2005).
4. Tang, J.-M. & Flatté, M. E. *Phys. Rev. Lett.* **92**, 047201 (2004).
5. Mahadevan, P., Zunger, A. & Sarma, D. D. *Phys. Rev. Lett.* **93**, 177201 (2004).
6. Wolf, S. A. *et al. Science* **294**, 1488–1495 (2001).
7. Yakunin, A. M. *et al. Phys. Rev. Lett.* **92**, 216806 (2004).
8. Yakunin, A. M. *et al. Phys. Rev. Lett.* **95**, 256402 (2005).
9. Eigler, D. M. & Schweizer, E. K. *Nature* **344**, 524–526 (1990).
10. Hirjibehedin, C. F., Lutz, C. P. & Heinrich, A. J. *Science* **312**, 1021–1024 (2006).

PROTEIN FOLDING

Inside the cage

R. John Ellis

Many newly synthesized bacterial proteins avoid aggregation by folding inside a chaperonin nanocage. Unexpectedly, it turns out that the cage's internal properties can be optimized to accelerate folding.

Proteins are the action molecules of life, but cells face a problem in making them. Proteins consist of amino acids joined into linear polypeptide chains by intracellular structures called ribosomes. Each chain folds into a compact shape whose surface properties determine the biological function unique to that protein. The information required to fold correctly resides in the sequence of amino acids, and this is determined by the ribosomes, which translate a messenger RNA copy of the gene for each chain. But to synthesize proteins rapidly enough, each mRNA molecule is translated by more than one ribosome at the same time. The result is that partly folded identical chains accumulate within touching distance of one another, raising the possibility that they will bind together to form non-functional aggregates.

This universal problem is combated by proteins called molecular chaperones, of which more than 50 families are known¹. One family is the chaperonins, the best-studied member being GroEL–GroES, found in bacteria. GroEL–GroES prevents aggregation by encapsulating each partly folded chain inside its cage structure, where the chain can continue to fold in isolation from similar chains^{2,3}. In earlier work, the laboratory of Ulrich Hartl found⁴ that a particular protein — bacterial Rubisco — folds three to four times faster inside the cage than it does in free solution under conditions where aggregation is minimized. Writing in *Cell*⁵, this lab now reports that both the size and surface charge of the cage are optimized to speed up the folding of several different types of chain.

GroEL and GroES bind to one another in the presence of the nucleotides ATP or ADP to create a large complex containing a cavity, termed a cage (or nanocage), at one end (Fig. 1). GroES acts as a removable lid to keep the protein chain inside the cage while it is folding. Inside this cage, the chain continues to fold until its hydrophobic amino acids, which cause aggregation, are buried within the correctly folded compact protein. This cage was initially termed an Anfinsen cage⁶ to indicate the assumption that the chain folds inside the cavity in a manner determined by its amino-acid sequence, just as a denatured protein refolds in the classical experiment of Christian Anfinsen. The new experiments⁵ show that this model is incomplete, because the folding rate of some proteins depends on the relative sizes of the folding chain and the cage, and on the inside surface properties of the cage. Thus, GroEL–GroES is more than an anti-aggregation device — it also enables some proteins to fold faster than they do outside the cage.

About 85 different proteins of the bacterium *Escherichia coli* are thought to require encapsulation inside GroEL–GroES to fold correctly. Of these, 60% are 30–50 kilodaltons in size, and only 14% are greater than 50 kDa in size⁷. The cage of GroEL–GroES measures 80 × 85 Å, sufficient in principle to house proteins up to 70 kDa. However, the available volume is somewhat less, owing to the presence of 23 amino acids at the end of each GroEL subunit. Removal of these 'tails' does not affect the basic mechanism, so they provide a way of changing the size of the cage. By removing the tail or extending it stepwise, Tang *et al.*⁵ altered the

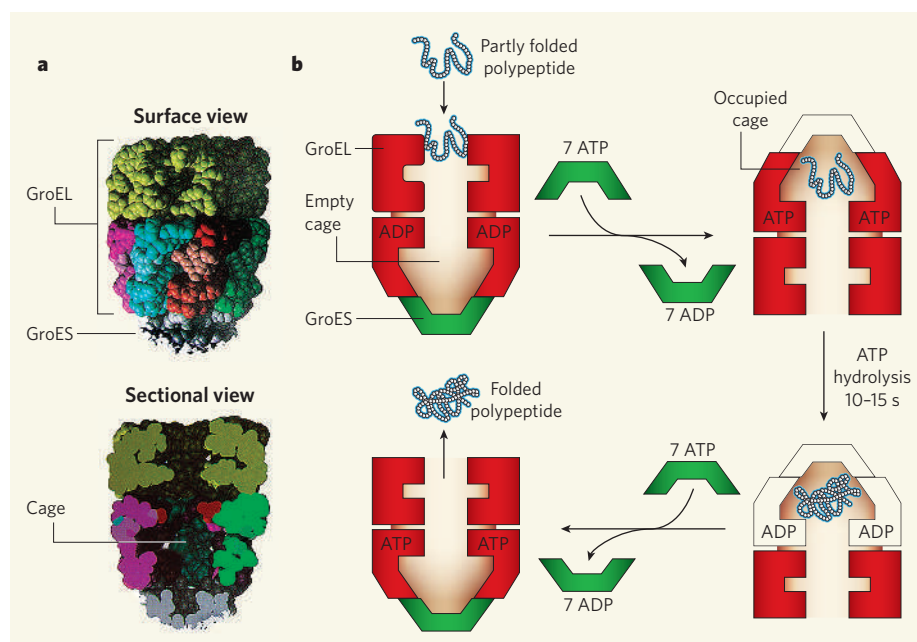


Figure 1 | The GroEL–GroES nanocage in action. **a**, Space-filling model of the GroEL–GroES complex. GroEL consists of two rings, each made of seven ATPase subunits arranged back to back, whereas GroES consists of a single dome-shaped ring of seven smaller subunits. The nucleotide-binding abilities of the two GroEL rings are mutually exclusive (that is, only one ring at a time can bind either ATP or ADP). In the presence of ATP or ADP, GroES binds to the same ring that binds the nucleotide, creating a cage at that end. (Reprinted from ref. 10.) **b**, A single, partly folded polypeptide chain released from the ribosome binds to hydrophobic residues exposed at the end of GroEL uncapped by GroES. Binding of GroES and ATP to that end triggers release of the chain into the newly created cage because GroES competes for binding sites with the partly folded chain. The chain has 10–15 seconds to fold inside this cage, because this is the time it takes for the ATP to be hydrolysed to ADP by the ring that contains the folding protein. The ADP then unbinds, allowing ATP to bind to the opposite ring. This binding triggers the release of GroES and of the folded polypeptide from the ring containing the cage. If the released chain has not internalized all its hydrophobic residues, it is likely to be rebound by the same ring. (Revised from ref. 1.)

volume of the cage in increments from + 4% to – 13%, and measured the effects of these changes on the rate of folding of four different proteins in the size range 33–50 kDa.

The folding rate of the smaller proteins rhodanese (33 kDa) and MetF (33 kDa) increased by up to two times when the volume of the cage was reduced by 4.4%, but reduction by 8.7% returned the rate to that found in the normal, wild-type chaperonin. Further reduction to 13% decreased the rate fivefold for both proteins. In the case of rhodanese, the spontaneous rate of folding outside the cage, under conditions where aggregation was minimal, was the same as inside the wild-type cage — a clear difference from Rubisco. For the larger substrates, maltose-binding protein (41 kDa) and bacterial Rubisco (50 kDa), either reducing or increasing the cage volume slowed the folding. The same results were observed when the experiment was repeated with a single-ring mutant of GroEL that cannot release GroES, indicating that the rate changes reflect effects of encapsulation rather than protein release into solution. The conclusion is that the cage has an optimal volume for folding rate, depending on the size of the encapsulated protein. So size matters. What about other factors?

The interior wall of the cage contains 189 negatively charged amino-acid residues but

only 147 positively charged residues. Tang *et al.*⁵ altered the overall net negative charge of 42 residues in steps down to zero by replacing some of them with neutral or positively charged residues. There were several effects on protein-folding rate. For example, reducing the net charge to zero abolished the folding of Rubisco but enhanced the folding rate of rhodanese by about 50%. The original idea⁶ that GroEL–GroES is an inert protective container is clearly incorrect — it is an active protective container, at least *in vitro*.

But does this activity have any biological relevance? By co-expressing the modified GroEL–GroES complexes with the same substrate proteins in intact cells of *E. coli*, Tang *et al.*⁵ show that reducing the ability to accelerate folding also reduces the amount of some proteins that fold correctly inside the cell. The size and surface properties of the cage represent an evolutionary compromise that helps the bacterial cell to produce functional proteins fast enough to survive in a competitive microbial world.

The effects of cage volume on folding rate are consistent with the predictions of a branch of macromolecular crowding theory called confinement⁸. Crowding describes the fact that the total concentration of macromolecules inside cells is so high that the thermodynamic



50 YEARS AGO

“Energy requirements of Europe”

— The report on Europe’s future needs of energy recently issued by the Organization for European Economic Co-operation contains a most interesting examination of the position... No doubt in time the competitive position of the coal industry may be affected, for after 1975 nuclear energy is expected to free the coal burnt in thermal power stations to an increasing extent... The share of oil in supplying the energy needs of Western Europe must also rise rapidly... The world’s reserves of oil will be adequate for this requirement, but large investments must be made by the oil companies in Western Europe as well as abroad... A potent factor in reducing the menacing gap between indigenous production and the overall requirements of energy resides in the still existing possibilities of saving energy at every stage, from its winning and conversion...to its final utilization in industry and in the home.

From *Nature* 28 July 1956.

100 YEARS AGO

“With wires and without” — Of the numerous achievements of which the electrical engineer can boast, telegraphy is the one of which he has greatest reason to be proud. If we combine with telegraphy the sister subject of telephony, there can be little doubt but that by the application of these two sciences he has effected a greater revolution in human affairs than by all successes in the way of heavy engineering. He may “electrify” our railways, especially the suburban lines, to the great advantage of both the travelling public and the shareholder, but he is still only doing for us in another way what the mechanical engineer has already accomplished. He may harness great waterfalls and transmit their power over hundreds of miles to localities at which it may be more easily utilised, but he is only saving Mahomet the trouble of going to the mountain... But with telegraphy he has given us something entirely new—an art which, while actually annihilating distance, virtually annihilates time.

From *Nature* 26 July 1906.

50 & 100 YEARS AGO

activities of these molecules are up to three orders of magnitude greater than in dilute solution. This dramatic effect results from the fact that 8–40% of the volume of the cell is physically occupied by the macromolecules themselves and is therefore unavailable to other molecules. This exclusion of part of the volume has energetic consequences — any process that reduces the excluded volume is favoured, because this reduction lowers the free energy of the system. Such processes include protein folding, because the folded chain is more compact than the partly folded chain. The size of the cage affects folding because it restricts the number of extended conformations that the chain can adopt as it folds (there is less room to swing the cat); and the net negative charge of the cage wall repels the folding chain because most GroEL substrates have a net negative charge. But excluded volume is also reduced by protein aggregation⁹, which brings us back to where we started. This effect of crowding on protein aggregation explains why cells require molecular chaperones.

It is a testament to the ingenuity of natural selection that the chaperonin cage not only combats aggregation caused by crowding outside the cage but also uses crowding to accelerate protein folding inside the cage. Nanoengineers trying to improve the yield of therapeutic proteins could profit from studying the tricks of the chaperonin nanocage. ■

R. John Ellis is in the Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK.

e-mail: jellis@bio.warwick.ac.uk

1. Ellis, R. J. in *Molecular Chaperones and Cell Signalling* (eds Henderson, B. & Pockley, G.) 3–21 (Cambridge Univ. Press, 2005).
2. Saibil, H. R. & Ranson, N. A. *Trends Biochem. Sci.* **27**, 627–632 (2002).
3. Fenton, W. A. & Horwich, A. L. *Q. Rev. Biophys.* **36**, 229–256 (2003).
4. Brinker, A. et al. *Cell* **107**, 223–233 (2001).
5. Tang, Y.-C. et al. *Cell* **125**, 903–914 (2006).
6. Ellis, R. J. *Fold. Des.* **1**, R9–R15 (1996).
7. Kerner, M. J. et al. *Cell* **122**, 209–220 (2005).
8. Minton, A. P. *J. Biol. Chem.* **276**, 10577–10580 (2001).
9. Ellis, R. J. & Minton, A. P. *Biol. Chem.* **387**, 485–497 (2006).
10. Xu, Z. H. et al. *Nature* **388**, 741–749 (1997).

PLANETARY SCIENCE

Titan's exotic weather

Caitlin A. Griffith

Titan is viewed as a sibling of Earth, as both bodies have rainy weather systems and landscapes formed by rivers. But as we study these similarities, Titan emerges as an intriguingly foreign world.

Images of Saturn's largest moon Titan, taken during the joint NASA/European Space Agency Cassini–Huygens mission, invoke a sense of familiarity: river channels meander downhill to damp lake-beds, where icy, rounded stones, resembling river cobbles, litter the ground^{1,2}; massive cumulus clouds form and quickly dissipate, suggestive of rain^{3–5}; and dark oval regions resemble lakes (Fig. 1). These features arise from Titan's unique similarity to Earth: both bodies cycle liquid between their surfaces and atmospheres, but in Titan's cool atmosphere (–179 °C at the surface) it is methane, rather than water, that exists as a gas, liquid and ice.

In this issue, two papers illuminate the nature of Titan's weather. One study, by Tokano *et al.*⁶ (page 432), forecasts light drizzle over much of Titan's surface for the next few years. This is consistent with the damp surface observed by the Huygens probe at its landing site (at 10° S latitude, 192° W longitude). But such drizzle is too delicate to create the fluvial features that cut through Titan's hard surface. The second study, by Hueso and Sánchez-Lavega⁷ (page 428), suggests how these river valleys may have formed. Models of Titan's storms predict torrential rain strong enough to carve the rugged washes. Yet no



Figure 1 | Alien landscape. An image of Titan's surface taken by the Huygens probe¹, which descended into Titan's atmosphere in January 2005. River channels flow from the lighter-coloured hills into a damp lake-bed, which is riddled with 'cobbles'.

storms were seen at the Huygens landing site, nor was there any indication of how storms might develop in the calm equatorial atmosphere of that area⁶.

Although Titan is tantalizingly similar to Earth, its atmosphere is ten times thicker, and much cooler, so that it takes longer than a Titan year (29.5 Earth years) for the atmosphere to respond radiatively to seasonal heat-

ing changes. Titan also lacks oceans — which are central to Earth's climate — and instead stores much of its condensable matter in its atmosphere. As a result, Titan's weather differs markedly from Earth's. Evidence for this is given by the location of Titan's large clouds, which presently reside either in a narrow band at 40° S latitude or within 30° of the south pole^{8–10}. In contrast, terrestrial clouds visit all latitudes (Fig. 2a). Yet they too exhibit striking latitudinal patterns — they occur preferentially near the Equator, but are scarce at latitudes of 15–35° N and S (regions that contain many deserts). Clouds also dominate the 40–60° northern and southern skies, covering northern Europe, for example.

These terrestrial cloud patterns are partly explained by a simple model of Earth's atmospheric circulation. Sunlight illuminates Earth most directly at the Equator, where air heats and rises to around 15 kilometres, expanding and then cooling, so that water vapour condenses to form tropical thunderstorms. The air is unable to rise further, so it spreads away from the Equator to higher latitudes (±15–35°). Here, the dried-out air gradually loses heat and sinks to the desert latitudes, where sunlit, parched surfaces warm it up. At latitudes of 40–60° N and S, atmospheric circulation also contributes to cloud formation. Here, cool, dense polar air meets and slides underneath warmer lower-latitude air, which ascends. As the warmer air rises, it expands and cools, so that water vapour within it condenses to form clouds.

So, does atmospheric circulation correlate with cloud formation on Titan? Recent models¹¹ suggest that it does. Titan's southern hemisphere is currently enjoying its long summer. Here, the Sun's warmth causes air to rise, particularly at 40° S, where clouds are observed. At around 70° S, low-latitude air converges with cooler polar air¹¹; clouds are also present in this area (Fig. 2b). The regions of ascending air, and so of cloud formation, are expected to switch to the north (to around 40° and 70° N) for Titan's northern summer. In the three-year period between northern and southern summers, updrafts occur at the equator, but their effect is predicted to be too weak to cause storms when averaged over all longitudes¹¹.

So far, so good, but what is the detailed nature of Titan's storms? Hueso and Sánchez-Lavega⁷ present models of the formation and evolution of Titan's storms with a resolution of 0.5 km. These models describe the small-scale structure of discrete clouds that are formed by updrafts of air. The authors find that small temperature perturbations of 0.5 °C and updrafts of 1 m s^{–1} initiate the formation of Titan's high clouds, and explain their altitudes and morphologies. These models also predict unexpectedly heavy rainfall, equivalent to severe terrestrial storms.

A picture emerges of the pole-to-pole migration of rainstorms^{7,11} that shapes river valleys and creates lakes. The leftover rivers

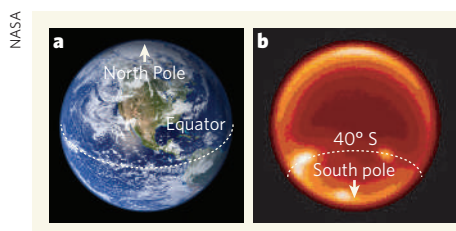


Figure 2 | Sibling atmospheres. **a**, Earth's clouds, although present at all latitudes, predominate in a band near the Equator and as large systems at high northern and southern latitudes. **b**, Ground-based images of Titan's clouds¹⁰ (bright features), which reside almost exclusively at 40° S latitude and near the south pole. Their hourly variations and discrete morphologies indicate a convective evolution typical of thunderstorms^{3,4}.

and metre-deep lakes evaporate during the dry seasons, into an atmosphere that can hold enough methane to cover Titan's surface to a depth of 4 m. However, it is still unclear how storms form in Titan's tropics. Perhaps atmospheric tides from Saturn¹² cause stronger local updrafts to occur during the equinox than are currently predicted. This effect, possibly in combination with others, such as solar warming of a porous surface¹³, may cause tropical storms at the equinox⁷.

Clouds can also form without strong updrafts. Such stratiform clouds develop, for example, over Earth's Arctic in the summer, as warm, low-latitude air flows over the cooler ice shelves. Tokano *et al.*⁶ find that stratiform clouds also exist on Titan. Measurements taken by the Huygens probe^{14,15} show that Titan has a complicated methane distribution over the tropics. Here, the abundance of methane is mysteriously pegged at around 80% humidity, assuming pure methane condensation, at altitudes of 8–16 km, and rises to 100% humidity at around 21 km. The authors suggest that below 16 km, where the clouds consist of liquid droplets, it is not pure methane that condenses; a liquid solution of nitrogen dissolved in methane condenses instead. The atmosphere is found to be saturated at altitudes of 8–16 km with respect to methane–nitrogen liquid condensation, and saturated at around 21–30 km for pure methane–ice condensation. These wet conditions indicate the existence of stratiform methane clouds at 21–30 km, as seen by Huygens, and tenuous methane–nitrogen clouds at 8–16 km, not yet detected. The highly stable conditions in Titan's tropics suggest that the resulting drizzle is typical of this region over most of Titan's year, except possibly near the equinox.

Currently, only Titan's atmospheric circulation and its extremely stable, methane-saturated conditions have been associated with cloud formation^{6,7}. The combined effects of its patchy wet surface, atmospheric tides, possible ice volcanoes and detailed seasonal variations remain unclear, as we have witnessed only one Titan season. Many questions remain unanswered — for example, we cannot yet explain

observations of extensive tropical dunes¹⁶, which require dry conditions, when long-term drizzle is the forecast. Even the source of atmospheric methane is a mystery.

If Cassini is still operational in 2010, perhaps it will witness the storms and flooding that create, lift and transport cobbles, while cutting drainage systems through a land that now languishes in a calm, eerie drizzle. Although clouds on both Titan and Earth may, on average, be married to their atmospheric circulations, Titan's weather is enticingly alien.

Caitlin A. Griffith is in the Department of Planetary Sciences, University of Arizona, Tucson, Arizona 85721, USA.
e-mail: griffith@lpl.arizona.edu

1. Tomasko, M. G. *et al.* *Nature* **438**, 765–778 (2005).
2. Elachi, C. *et al.* *Science* **308**, 970–974 (2005).
3. Griffith, C. A., Hall, J. L. & Geballe, T. R. *Science* **290**, 509–513 (2000).
4. Porco, C. C. *et al.* *Nature* **434**, 159–168 (2005).
5. Griffith, C. A. *et al.* *Science* **310**, 474–477 (2005).
6. Tokano, T. *et al.* *Nature* **442**, 432–435 (2006).
7. Hueso, R. & Sánchez-Lavega, A. *Nature* **442**, 428–431 (2006).
8. Brown, M. E., Bouchez, A. H. & Griffith, C. A. *Nature* **420**, 795–797 (2002).
9. Roe, H. G., de Pater, I., Macintosh, B. A. & McKay, C. P. *Astrophys. J.* **581**, 1399–1406 (2002).
10. Roe, H. G., Bouchez, A. H., Trujillo, C. A., Schaller, E. L. & Brown, M. E. *Astrophys. J.* **618**, L49–L52 (2005).
11. Rannou, P., Montmessin, F., Houdin, F. & Lebonnois, S. *Science* **311**, 201–205 (2006).
12. Tokano, T. *Icarus* **158**, 499–515 (2002).
13. Tokano, T. *Icarus* **173**, 222–242 (2005).
14. Niemann, H. B. *et al.* *Nature* **438**, 779–784 (2005).
15. Fulchignoni, M. *et al.* *Nature* **438**, 785–791 (2005).
16. Lorenz, R. *et al.* *Science* **312**, 724–727 (2006).

DEVELOPMENTAL BIOLOGY

The hole picture

Keith Mostov and Fernando Martin-Belmonte

What's the best way to make a tube? Roll up a sheet? Hollow out a solid rod? Some innovative movies show how the problem is tackled during the development of blood vessels in embryos.

In simple organisms, only one or a few cells thick, diffusion brings nutrients and oxygen to individual cells and removes their waste. But in larger animals, this process is no longer sufficient, so transport tubes evolved to allow materials to move quickly throughout the organism. Such transport systems include blood vessels and the tubes that make up the digestive, respiratory and other organ systems. Generally these tubes are lined by a single layer of cells known as epithelial cells, but the inner surface of blood vessels are coated with modified epithelial cells called endothelial cells. Tubes must have a continuous hollow space, or lumen, in the centre to allow free passage of materials. A century-old question is how lumina form in tubes during development. Kamei *et al.* (page 453 of this issue)¹ have watched living embryos to discover how lumina are made during the development of fish.

Lumen formation has been examined previously using static images of endothelial and epithelial cells in culture. A commonly observed feature in these pictures is the formation of large² vacuoles — membrane-bounded compartments — in the cell (Fig. 1, overleaf). These are thought to be formed by a process called endocytosis, in which cells suck in a small portion of their outer, plasma membrane to form a pocket and then pinch it off to create an internal 'bubble' containing some of the extracellular fluid. It had been imagined that these vacuoles fuse together and with the plasma membrane to produce lumina between cells, although the evidence, based on static images, was not definitive.

Now Kamei *et al.* have filmed this process, first in an endothelial culture system and then in the developing blood vessels of fish embryos — which fortunately are transparent. In both cases the movies show actual fusion of the vacuoles with the plasma membrane to create lumina between cells (Fig. 1). In essence, the cells first create tiny intracellular lumina by sipping up extracellular fluid into endocytic vacuoles. These then fuse with the plasma membrane, thereby combining many small intracellular lumina into a single extracellular lumen between the cells that will line the nascent tube.

These movies of lumen creation are astounding: if a picture is worth a thousand words, the power of these films to answer a long-standing question is certainly worth a million. Kamei *et al.* exploit this power to show how developing lumina extend from one cell to another to lengthen the tube. To do so, they injected nanometre-sized fluorescent beads (quantum dots) into the blood circulation of the fish embryo. The quantum dots appeared first in a large blood vessel, the dorsal aorta. Then, as a new, smaller vessel began to bud off, the quantum dots emerged in the lumen of a previously unlabelled vacuole in an endothelial cell adjoining the aorta. Quantum dots showed up successively in a second and then a third neighbouring cell, as each cell's intracellular vacuole fused with its plasma membrane to extend the lumen of the nascent endothelial tube.

Many other epithelial organs face the problem of how to create a lumen. Superficially, at least, there is a multitude of ways by which

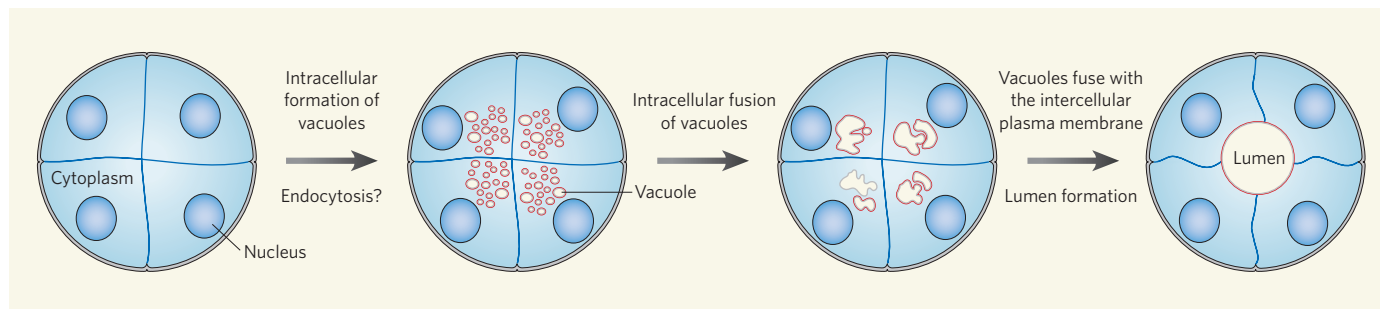


Figure 1 | Making spaces. Kamei *et al.*¹ show that lumina in vascular tubes are formed by the fusion of intracellular vacuoles. Vacuoles form inside the endothelial cells, apparently by endocytosis. These vacuoles may fuse with each other to create larger vacuoles. The vacuoles then fuse with the plasma membrane, to create lumina between the cells.

such tubes form in different organs, but there should be common mechanisms underlying this diversity^{3,4}. In some cases, especially larger tubes, cells in the middle may undergo programmed cell death to hollow out a lumen⁵. Large vacuoles are not usually seen inside epithelial cells in the process of forming lumina and tubes, although they do occur in certain situations⁶. It may be that in most circumstances, much smaller vacuoles or related membrane-bounded 'vesicles' are used to build the luminal surface of the epithelial cell⁷. These vesicles may be part of the endocytic pathway or of the pathway by which newly made membrane is normally delivered to the cell surface, or both.

Understanding the morphogenesis of multicellular structures, such as tubes, therefore

requires analysis of the molecular mechanisms behind the movement and formation of membranes. Such membrane trafficking depends on a special class of fats or lipids, known as phosphatidylinositides, which are components of the greasy bilayer that make up all cellular membranes. One such lipid, phosphatidylinositol 3-phosphate, is characteristically found mainly on the membrane-bounded compartments of the endocytic system. Related phosphatidylinositides tend to occupy specific locations in the cell, leading to the idea that various phosphatidylinositides could determine the identity of different membrane compartments of the cell⁸. Perhaps the phosphatidylinositide composition of the plasma membrane facing the lumen of tubes holds the key to its assembly and identity. ■

Keith Mostov and Fernando Martin-Belmonte are in the Departments of Anatomy, and Biochemistry and Biophysics, University of California, San Francisco, California 94143-2140, USA.
e-mail: Keith.Mostov@ucsf.edu

1. Kamei, M. *et al.* *Nature* **442**, 453–456 (2006).
2. Davis, G. E. & Bayless, K. J. *Microcirculation* **10**, 27–44 (2003).
3. Lubarsky, B. & Krasnow, M. A. *Cell* **112**, 19–28 (2003).
4. Hogan, B. L. & Kolodziej, P. A. *Nature Rev. Genet.* **3**, 513–523 (2002).
5. Debnath, J. & Brugge, J. S. *Nature Rev. Cancer* **5**, 675–688 (2005).
6. Vega-Salas, D. E., Salas, P. J. & Rodríguez-Boulán, E. *J. Cell Biol.* **104**, 1249–1259 (1987).
7. O'Brien, L. E., Zegers, M. M. P. & Mostov, K. E. *Nature Rev. Mol. Cell Biol.* **3**, 531–537 (2002).
8. De Matteis, M., Godi, A. & Corda, D. *Curr. Opin. Cell Biol.* **14**, 434–447 (2002).

ASTRONOMY

Revealing flares

J. Anthony Tyson

Transient bursts of cosmic light provide a unique window on what's going on in the distant Universe. But similar bursts closer to home may be muddying the view, and hopes rest on a new tool to resolve things.

Whenever we look at the Universe afresh, we are reminded of its inexhaustible richness. Although much of the sky seems immutable, scanning the heavens for change can teach us much about this richness and its underlying physics — observations of rare stellar explosions are increasingly the key to unlocking the Universe's secrets. Writing in *The Astrophysical Journal*¹, Kulkarni and Rau draw attention to a complication in observing transient phenomena from the far Universe: if astronomers wish to follow such far-off bursts of light, they must first take a foreground 'fog' of flaring stars in our own Galaxy into account.

Known as time-domain astronomy, the study of transient cosmic phenomena has assumed a prominent role in astronomical discoveries of the past few decades. Daily fluctuations in the brightness of 'quasi-stellar objects', which can outshine a hundred

galaxies, indicate they have supermassive black holes at their cores. Stellar explosions known as supernovae, which last only a few weeks, but which are visible across much of the Universe, have recently provided evidence for a previously unknown 'dark energy' that is accelerating the expansion of the Universe. And energetic events lasting minutes, or less, occur just where one might expect to discover extreme physical phenomena, such as the creation of a black hole. Yet transient events represent a relatively little-explored aspect of the distant Universe, partly because far-off transient objects are, by their nature, among the most difficult cosmic phenomena to observe.

The tools available to astronomers are an impediment to observing rapid change. Modern large-scale telescopes have a light-gathering power several million times that of the naked eye. But, remarkable as they are, they

have been designed to look deeply only at very small parts of the sky at any one time. They are thus unlikely to be pointing precisely the right way to catch a transient event 'in the act'. Their small field of view also implies an impractically large number of separate observations to map the entire sky to the distances required.

Instead, astronomers' all-sky maps are made with small telescopes that are restricted in the depth and detail they can achieve. Their limited light-gathering capability means that a single map takes years to complete — so it is nearly impossible to detect rapid changing events with these telescopes, too. Most of what we know about transient events we have discovered serendipitously, and we are almost certainly missing most of what is changing in the Universe.

Investigations of transients at optical wavelengths naturally start with the brightest events. In 1999, a spectacularly bright and rapid flare was shown to be associated with a burst of high-energy γ -rays coming from the distant Universe². More recently, a small optical patrol telescope, triggered by the Burst Alert Telescope on board NASA's Swift satellite — which was specially designed to look for γ -ray bursts — followed an associated burst of light at optical wavelengths that lasted for 4 minutes (ref. 3).

Despite these successes, our monitoring and knowledge of transient optical sources in the

cosmos are in their infancy^{4,5}. The detection of transient emission is important because it provides a window on diverse astrophysical objects, from variable stars and stellar explosions to the mergers of compact stellar remnants. Even more exciting is the potential for discovering new, unanticipated phenomena. The intriguing initial data hint at much more to come: imagine what we might find if we could monitor the sky for optical bursts just minutes long, and a million times fainter than those currently spotted.

By using larger, 4-metre-aperture telescopes to peer deeper into the distant Universe while still surveying a large volume, a few short-lived, faint optical bursts without apparent precursor objects have already been spotted. The Deep Lens Survey, for example, which searched an area of sky of 20 square degrees using the National Science Foundation's sister Blanco and Mayall telescopes in Chile and Arizona, found three blue flares with lifetimes shorter than 1,000 seconds. Follow-up spectroscopy, however, revealed one of these to be a flaring faint red dwarf star in our Galaxy⁶.

Using the twin 10-metre-aperture Keck telescopes on the summit of Mauna Kea, on Hawaii, Kulkarni and Rau¹ have obtained spectra of the remaining two mysterious flaring objects. Their conclusion is that they too are blue flares of red dwarf stars (Fig. 1). Extrapolating to the next generation of wide-field surveys, which will probe deeper into the distant Universe for transient events, the implication is that we will be peering through a foreground confusion of perhaps 7,000

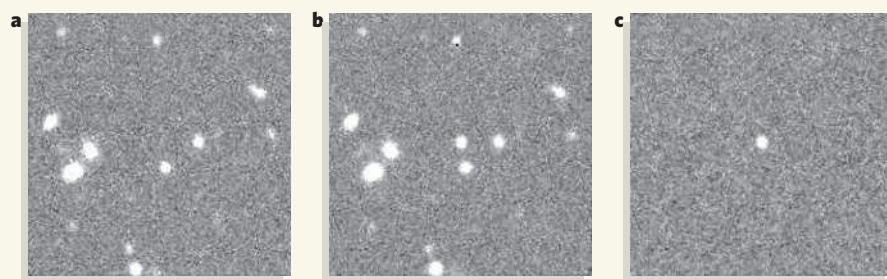


Figure 1 | Optical flash: cosmic or local? A short-lived optical flare is caught on camera in this sequence of Deep Lens Survey image mosaics taken with the Blanco telescope at the Cerro Tololo Interamerican Observatory in Chile. **a**, **b**, Images taken before and during the flare. **c**, Difference between images in **a** and **b**; the flaring object alone is seen. Kulkarni and Rau's spectroscopic results¹ reveal a faint red dwarf star in our Galaxy at that position. Such flaring foreground stars could confound searches for faint bursts of light from the distant Universe, unless images at wavelengths in the near infrared are also obtained as part of future surveys. (Image courtesy of A. Becker *et al.*⁵)

flaring dwarf stars in our Galaxy each year.

Help might be at hand. Taking advantage of advances in scientific computing that allow the fast processing of vastly increased quantities of data, as well as breakthroughs in large-scale optics and microelectronics, the Large Synoptic Survey Telescope (LSST)⁷ has been designed to overcome many of these difficulties. It will open up the time domain in astronomy by going wide, fast and deep, observing the entire visible sky 2,000 times in six colours, and producing up to 30,000 gigabytes of data every night. Cosmic cartography will become cosmic cinematography.

The LSST will be capable of separating distant high-energy events from blue flares on nearby stars by identifying the host star, which

would emit light at a near-infrared wavelength. More generally, the thousandfold increase in breadth and depth of view that can be surveyed during any one period will virtually guarantee the discovery of new transient phenomena. A new view of our Universe is in sight.

J. Anthony Tyson is in the Department of Physics, University of California, Davis, California 95616, USA.

e-mail: tyson@physics.ucdavis.edu

1. Kulkarni, S. R. & Rau, A. *Astrophys. J.* **644**, L63–L66 (2006).
2. Akerlof, C. *et al.* *Nature* **398**, 400–402 (1999).
3. Wozniak, P. R. *et al.* *Astrophys. J.* **642**, L99–L102 (2006).
4. Paczynski, B. *Publ. Astron. Soc. Pacif.* **112**, 1281–1283 (2000).
5. Colgate, S. A., Moore, E. P. & Carlson, R. A. *Publ. Astron. Soc. Pacif.* **87**, 565–575 (1975).
6. Becker, A. C. *et al.* *Astrophys. J.* **611**, 418–433 (2004).
7. www.lsst.org

DEVELOPMENTAL NEUROBIOLOGY

A destructive switch for neurons

Peter K. Jackson

In the developing nervous system, tremendous multiplication and diversification of cells elaborate the exquisite pattern of the brain. But how do cells shift from early proliferation to assume their mature states?

Control of cell division in the developing nervous system is a highly orchestrated process that sets up the patterns for the extended structure of the brain. In the embryonic brain, the division of neuroblasts — the precursors of neurons — occurs in specific proliferative zones. From these zones, neuroblasts undergo multiple rounds of cell division (mitosis) while migrating considerable distances, producing clusters of neuronal precursor cells. At specific times and sites in the developing brain, neuroblasts stop dividing and mature into neurons. The post-mitotic cells extend axonal processes, the elongated nerve fibres that form synapse junctions to communicate with other neurons. Although the details are not well understood, specific extracellular signals in

the developing brain seem to provide the spatial cues that position the birth of the mature neuron. There also seem to be intrinsic timing mechanisms within neuroblasts that stop the cell-division cycle and trigger the activation of the programme that extends the axon.

In this issue, Iavarone and colleagues¹ (Lasorella *et al.*, page 471) provide a crucial insight into this triggering mechanism. The authors show that a well-known regulator of exit from mitosis — the anaphase-promoting complex — directly modifies and initiates the destruction of Id2, a regulator of gene expression. This protein inhibits a protein complex called E12–E47 that initiates the expression of neuron-specific genes². The authors also show

that the activation of E47, following the destruction of Id2, may turn on genes that prepare the axon for later stages of neuronal maturation. Together, these findings open a new link between the mechanism that causes cells to cease division and the activation of the gene-expression programme that directs neuronal function (Fig. 1, overleaf).

The implications of this work extend well beyond the development of the nervous system, as the E12 and E47 proteins are members of the basic helix-loop-helix (bHLH) DNA-binding proteins that drive cellular differentiation in many cell lineages². These proteins function as heterodimers, where a tissue-specific bHLH, such as the E47 protein, pairs with the ubiquitous bHLH subunit E12 to activate a cell-type-specific gene-expression programme.

A family of inhibitors called 'inhibitor of differentiation' (Id) proteins also contain the bHLH domain, but lack the domain required for DNA binding. So the Id proteins bind to bHLH factors and quench their ability to activate transcription (gene expression). In several cell lineages, inhibition of the bHLH-driven differentiation programme by Id proteins maintains cells in the unspecialized, dividing state. Thus, a change in the balance between

Id proteins and bHLH transcriptional activity may be at the heart of the transition from cell division to differentiation in many tissues. But what triggers that transition has not been clear. The work by Iavarone and colleagues¹ suggests that destruction (proteolysis) of the Id2 protein may be the key. So, what's the destroyer, and how is it tied to the cell-division cycle?

In cell division, the periodic accumulation and destruction of proteins called mitotic cyclins forms the central clock directing cell division. When cyclin accumulates to sufficient levels, it triggers the activation of the central events in mitosis, including the segregation of chromosomes into daughter cells and the exit from the cell cycle. At the heart of the mechanism for mitotic exit is the anaphase-promoting complex/cyclosome (APC/C)³. This complex belongs to a large class of enzymes that tag specific proteins with a label that targets them for degradation⁴. The APC/C complex directs the destruction of cyclin and other important mitotic regulators, thereby initiating the mitotic exit. For APC/C to cause exit from the division cycle, it requires a specific activator called Cdh1. What was less clear was whether APC/C^{Cdh1} mostly eliminates general cell-division regulators, such as cyclins, or whether it has more specific targets that inhibit cell differentiation after mitosis, such as Id2.

Id2 is expressed in neuroblasts of several neural lineages, suggesting that it has a central role in neural development². To examine what this might be, Iavarone and colleagues purified proteins that associate with Id2 in embryonic mouse brains. Surprisingly, they found components of APC/C. Previous studies had identified substantial amounts of APC/C^{Cdh1} in post-mitotic neurons in the adult brain⁵, but this is the first demonstration that it binds a functionally pivotal target in these cells.

A look at the sequence of Id2 turned up a region known as the destruction box that is required for APC/C to bind to its substrates⁶ and to mark that protein for destruction³. The authors show that Id2 is indeed degraded in mouse embryonic neuroblasts after the cells exited mitosis, and that this degradation requires APC/C^{Cdh1} and an intact destruction box. It will be of interest to see whether APC/C and Id2 destruction have a role in adult, damaged or regenerating neural tissues.

Inhibition of Cdh1 causes overly elongated axons⁷, suggesting that Cdh1 triggers destruction of a regulator of axon growth. This regulator was unknown, but Id2 was a strong candidate. Iavarone and colleagues¹ now show that inactivation of Cdh1 in post-mitotic neurons stops the destruction of Id2. Also, expression of an Id2 mutant with a defective destruction box (which is therefore not broken down) stimulated a prolonged axon outgrowth in cerebellar granule neurons, cortical neurons and neuroblast cell lines. So, Id2 is clearly a central factor that is destroyed in post-mitotic neurons to restrain axon growth

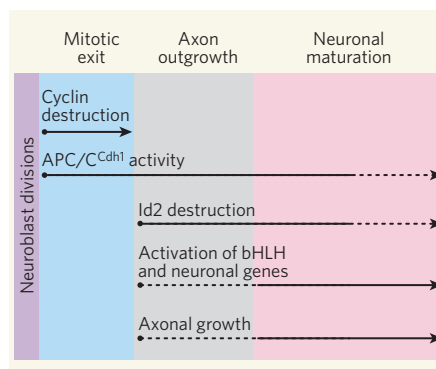


Figure 1 | Protein destruction in post-mitotic neurons. The work by Iavarone and colleagues¹ suggests that following the final neuroblast cell division (mitosis), cyclins and other mitotic regulators are marked for destruction by the Cdh1 form of the anaphase-promoting complex/cyclosome (APC/C^{Cdh1}). In the ensuing post-mitotic period, destruction of the Id2 transcriptional regulator allows activation of its target, the E47–E12 transcriptional regulator. Transcription of crucial neuronal genes follows and the gene products may both promote and eventually limit axon growth, but also promote later events in neuronal maturation.

— the next step in neuronal maturation. Another possible APC/C substrate, the transcriptional regulator SnoN, is also a target of Cdh1 in neurons⁸, but the relative importance of SnoN and Id2 destruction will require further study.

If blocking Id2 destruction causes pro-

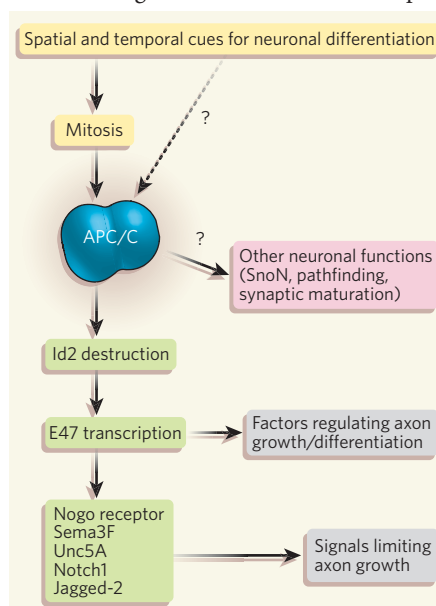


Figure 2 | A model for destruction of inhibitors of neuronal differentiation. Spatial and temporal cues may initiate mitosis in neuroblasts. Activation of the APC/C in mitosis triggers destruction of cyclin, to allow mitotic exit, and of Id2, to allow activation of the neuron-specific E47 transcription factor. Activation of E47 promotes the transcriptional activation of several factors that limit axon growth (Nogo receptor, Sema3F, Unc5A, Notch1, Jagged-2), but possibly other factors that regulate axon growth and differentiation.

longed axon outgrowth, what of the E47 transcription factor that Id2 inhibits? Does E47 reverse axon growth? Iavarone and colleagues confirm that overexpressing E47 in neurons did indeed reverse the ability of Cdh1 inactivation to prolong axon outgrowth. So the E47 transcriptional response seems to be a major determinant of the programme limiting axon outgrowth (Fig. 2). The authors identify a list of candidate genes that might link E47 to axon growth control, including — provocatively — several genes that are implicated in the inhibitory and repellent signalling for axon growth. These include a secreted molecule (Sema3F), a cell surface receptor ligand (Jagged-2), and some neuronal receptors (Nogo, Unc5A and Notch1).

These results suggest a scenario of neuronal differentiation such that, as Id2 is being destroyed and the axon is extending, the cells prepare for axon growth inhibition by the transcriptional activation of an inhibitory programme (Fig. 2). One of the candidates for that inhibitory programme would be the myelin-inhibitory signals, which block axon outgrowth. Indeed, Iavarone *et al.* show that the Id2–E47 balance modulates this inhibitory programme.

This study¹ has strong implications for both neural biology and differentiation control more generally. In neurons, the targets of E47 may shed considerable light on the regulation of axon outgrowth. Other bHLH factors, including the NeuroD regulator, may be regulated by Id2. Whether there are other transcriptional regulatory programmes under Id2 and/or APC/C control is also still unclear, although APC/C has been linked to axon target finding and synapse maturation^{9,10}.

Although APC/C-dependent destruction of Id2 is crucial in the transition to the post-mitotic programme, the specific timing and spatial signals that may trigger that decision are not well understood. There is considerable reason to suspect that polarity signals received by the mitotic cell might determine where and when the axon extends, possibly through regulating the activation of the APC/C towards substrates such as Id2. These studies linking cell-cycle exit, polarity and the differentiation of neurons may indeed have considerable extensions.

Peter K. Jackson is in the Department of Pathology, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, California 94305, and at Genentech Inc., 1 DNA Way, San Francisco, California 94080, USA. e-mail: pjackson@gene.com

- Lasorella, A. *et al.* *Nature* **442**, 471–474 (2006).
- Yokota, Y. *Oncogene* **20**, 8290–8298 (2001).
- Peters, J. M. *Mol. Cell* **9**, 931–943 (2002).
- Jackson, P. K. *et al.* *Trends Cell Biol.* **10**, 429–439 (2000).
- Gieffers, C. *et al.* *Proc. Natl Acad. Sci. USA* **96**, 11317–11322 (1999).
- Yamano, H. *et al.* *Mol. Cell* **16**, 137–147 (2004).
- Konishi, Y. *et al.* *Science* **303**, 1026–1030 (2004).
- Stegmüller, J. *et al.* *Neuron* **50**, 389–400 (2006).
- Juo, P. & Kaplan, J. M. *Curr. Biol.* **14**, 2057–2062 (2004).
- van Roessel, P. *et al.* *Cell* **119**, 707–718 (2004).

**Cover illustration**

Bubbles moving in a serpentine microchannel mix red and blue streams into purple. (Courtesy of G. M. Whitesides and M. Fuerstman.)

Editor, Nature

Philip Campbell

Insights Publisher

Sarah Greaves

Insights Editor

Karl Ziemelis

Production Editor

Sarah Archibald

Senior Art Editor

Martin Harrison

Art Editor

Nik Spencer

Layouts

Jane Walker

Sponsorship

Claire Hines
Claudia Banks

Production

Susan Gray

Marketing

Robin Black

Editorial Assistant

Laura Shaw

LAB ON A CHIP

The ability to perform laboratory operations on a small scale using miniaturized (lab-on-a-chip) devices is very appealing. Small volumes reduce the time taken to synthesize and analyse a product; the unique behaviour of liquids at the microscale allows greater control of molecular concentrations and interactions; and reagent costs and the amount of chemical waste can be much reduced. Compact devices also allow samples to be analysed at the point of need rather than a centralized laboratory.

Initially, however, pioneers of the field asked in *Chimia* whether their ideas about miniaturization would be “next century’s technology or just a fashionable craze”. The advantages are compelling, but designing and making devices of reduced size that operate effectively is challenging. The pioneers recognized the huge financial input and research effort needed to realize the full potential of the concept.

Now, well into that next century, it is clear that labs on chips are here to stay. Physicists and engineers are creating exciting functionality, and are starting to construct highly integrated compact devices. Chemists are using such tools to synthesize new molecules and materials, and biologists are using them to study complex cellular processes. Furthermore, labs on chips offer point-of-care diagnostic abilities that could revolutionize medicine. Such devices may find uses in other areas, including a range of industrial applications and environmental monitoring. Commercial exploitation has been slow, but is gaining pace, with some products now on the market. A technology for this century? The signs are looking good.

In this Insight, we present a collection of topical Reviews that discuss the history, design, application and future of lab-on-a-chip technologies, focusing on microfluidic flow devices. We hope you enjoy it.

We are pleased to acknowledge the financial support of Agilent Technologies in producing this Insight. As always, *Nature* carries sole responsibility for editorial content and peer review.

Rosamund Daw, Senior Editor

Joshua Finkelstein, Associate Editor

OVERVIEW

368 The origins and the future of microfluidics

G. M. Whitesides

REVIEWS

374 Scaling and the design of miniaturized chemical-analysis systems

D. Janasek, J. Franzke & A. Manz

381 Developing optofluidic technology through the fusion of microfluidics and optics

D. Psaltis, S. R. Quake & C. Yang

387 Future lab-on-a-chip technologies for interrogating individual molecules

H. Craighead

394 Control and detection of chemical reactions in microfluidic systems

A. J. deMello

403 Cells on chips

J. El-Ali, P. K. Sorger & K. F. Jensen

412 Microfluidic diagnostic technologies for global public health

P. Yager, T. Edwards, E. Fu, K. Helton, K. Nelson, M. R. Tam & B. H. Weigl

nature
insight

The origins and the future of microfluidics

George M. Whitesides¹

The manipulation of fluids in channels with dimensions of tens of micrometres — microfluidics — has emerged as a distinct new field. Microfluidics has the potential to influence subject areas from chemical synthesis and biological analysis to optics and information technology. But the field is still at an early stage of development. Even as the basic science and technological demonstrations develop, other problems must be addressed: choosing and focusing on initial applications, and developing strategies to complete the cycle of development, including commercialization. The solutions to these problems will require imagination and ingenuity.

What is microfluidics? It is the science and technology of systems that process or manipulate small (10^{-9} to 10^{-18} litres) amounts of fluids, using channels with dimensions of tens to hundreds of micrometres. The first applications of microfluidic technologies have been in analysis, for which they offer a number of useful capabilities: the ability to use very small quantities of samples and reagents, and to carry out separations and detections with high resolution and sensitivity; low cost; short times for analysis; and small footprints for the analytical devices¹. Microfluidics exploits both its most obvious characteristic — small size — and less obvious characteristics of fluids in microchannels, such as laminar flow. It offers fundamentally new capabilities in the control of concentrations of molecules in space and time.

As a technology, microfluidics seems almost too good to be true: it offers so many advantages and so few disadvantages (at least in its major applications in analysis). But it has not yet become widely used. Why not? Why is every biochemistry laboratory not littered with 'labs on chips'? Why does every patient not monitor his or her condition using microfluidic home-test systems? The answers are not yet clear. I am convinced that microfluidic technology will become a major theme in the analysis, and perhaps synthesis, of molecules: the advantages it offers are too compelling to let pass. Having said that, the answers to questions concerning the time and circumstances required for microfluidics to develop into a major new technology are important not just for this field, but also for other new technologies struggling to make it into the big time.

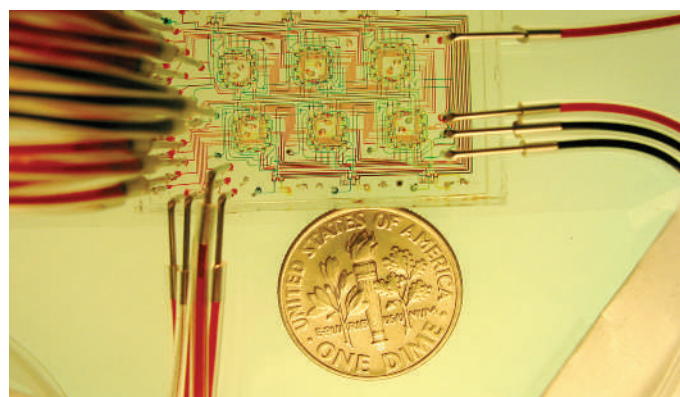


Figure 1 | A microfluidic chemostat. Microfluidic devices — here, a microfluidic chemostat used to study the growth of microbial populations — now routinely incorporate intricate plumbing. This device includes a high density of pneumatic valves. The colours are dyes introduced to trace the channels. (Image reproduced, with permission, from ref. 65.)

The field of microfluidics has four parents: molecular analysis, bio defence, molecular biology and microelectronics. First came analysis. The distant origins of microfluidics lie in microanalytical methods — gas-phase chromatography (GPC), high-pressure liquid chromatography (HPLC) and capillary electrophoresis (CE) — which, in capillary format, revolutionized chemical analysis. These methods (combined with the power of the laser in optical detection) made it possible to simultaneously achieve high sensitivity and high resolution using very small amounts of sample. With the successes of these microanalytical methods, it seemed obvious to develop new, more compact and more versatile formats for them, and to look for other applications of microscale methods in chemistry and biochemistry.

A second, different, motivation for the development of microfluidic systems came with the realization — after the end of the cold war — that chemical and biological weapons posed major military and terrorist threats. To counter these threats, the Defense Advanced Research Projects Agency (DARPA) of the US Department of Defense supported a series of programmes in the 1990s aimed at developing field-deployable microfluidic systems designed to serve as detectors for chemical and biological threats. These programmes were the main stimulus for the rapid growth of academic microfluidic technology.

The third motivational force came from the field of molecular biology. The explosion of genomics in the 1980s, followed by the advent of other areas of microanalysis related to molecular biology, such as high-throughput DNA sequencing, required analytical methods with much greater throughput, and higher sensitivity and resolution than had previously been contemplated in biology. Microfluidics offered approaches to overcome these problems.

The fourth contribution was from microelectronics. The original hope of microfluidics was that photolithography and associated technologies that had been so successful in silicon microelectronics, and in microelectromechanical systems (MEMS), would be directly applicable to microfluidics. Some of the earliest work in fluidic microsystems did, in fact, use silicon and glass, but these materials have largely been displaced by plastics. For analyses of biological samples in water, devices fabricated in glass and silicon are usually unnecessary or inappropriate. Silicon, in particular, is expensive, and opaque to visible and ultraviolet light, so cannot be used with conventional optical methods of detection. It is easier to fabricate the components required for microanalytical systems — especially pumps and valves — in elastomers than in rigid materials. Neither glass nor silicon has all the properties (especially permeability to gases) required for work with living mammalian cells.

¹Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA.

Thus, microfluidic devices have not developed as clones of silicon microelectronic devices. Much of the exploratory research in microfluidic systems has been carried out in a polymer — poly(dimethylsiloxane), or PDMS — the properties of which are entirely distinct from those of silicon^{2,3}. PDMS is an optically transparent, soft elastomer. Whether the microfluidic devices that ultimately become widely used will use PDMS or one of the engineering polymers (such as polycarbonate or polyolefin) remains to be seen. The ease with which new concepts can be tested in PDMS, however, and its ability to support certain very useful components (such as pneumatic valves), have made it the key material for exploratory research and research engineering at the early stages of development. Microelectronic technologies have, however, been indispensable for the development of microfluidics, and as the field has developed, glass, steel and silicon have again emerged as materials with which to build specialized systems that require chemical and thermal stability. The mechanical stability of silicon and glass are also useful in the nascent field of nanofluidics (the study of fluids in channels with nanometre-scale — ideally less than 50 nm — dimensions), in which channels with rigid walls can be useful^{4,5}.

Microfluidics has seen the rapid development of new methods of fabrication, and of the components — the microchannels that serve as pipes, and other structures that form valves^{6,7}, mixers^{8–10} and pumps¹¹ — that are essential elements of microchemical ‘factories’ on a chip. However, its impact on science has not yet been revolutionary. Revolutions in technology require both a broad range of different types of component and subsystem, and their integration into complete, functional systems. The field of microfluidics is in early adolescence, and still lacks both these essential requirements, in addition to the integration of components into systems that can be used by non-experts. As a field, it is a combination of unlimited promise, pimples and incomplete commitment. This is a very exciting time for the field, but we still do not know exactly what it will be when it grows up.

The present

A microfluidic system must have a series of generic components: a method of introducing reagents and samples (probably as fluids, although ideally with the option to use powders); methods for moving these fluids around on the chip, and for combining and mixing them; and various other devices (such as detectors for most microanalytical work, and components for purification of products for systems used in synthesis). The field has, so far, centred on demonstrating concepts for these components. Two particularly important contributions have been the development of soft lithography in PDMS as a method for fabricating prototype devices¹²; and the development of a simple method of fabricating pneumatically activated valves, mixers and pumps on the basis of soft-lithographic procedures¹³. These methods have made it possible to fabricate prototype devices that test new ideas in a time period much shorter (typically less than 2 days from design to working device) than that which could be achieved using silicon technology (typically, for non-specialists, a month or more). Quake’s pneumatic valves are particularly important as components that have enabled the design and examination of complicated devices, and these have opened up a number of areas of application (Fig. 1). ‘Quake valves’ use the restriction of a fluidic channel by an adjacent channel under pressure; their operation depends on the fact that PDMS is an elastomer, and no corresponding devices exist (or can exist) in rigid materials such as silicon and glass (or rigid engineering polymers such as polycarbonate).

Together with new methods of fabrication, microfluidics has been able to exploit certain fundamental differences between the physical properties of fluids moving in large channels and those travelling through micrometre-scale channels^{14–16}. Janasek *et al.* describe scaling relations that relate (or differentiate) macroscopic and microfluidic systems, with special emphasis on lab-on-a-chip devices (see page 374). Of these differences, the most important is turbulence (or its absence: laminar flow). On large scales, fluids mix convectively: for example, the mixing of milk when it is swirled into coffee, or smoke, leaving a chimney, with air. This type of mixing reflects the fact that in macroscopic fluids, inertia is often

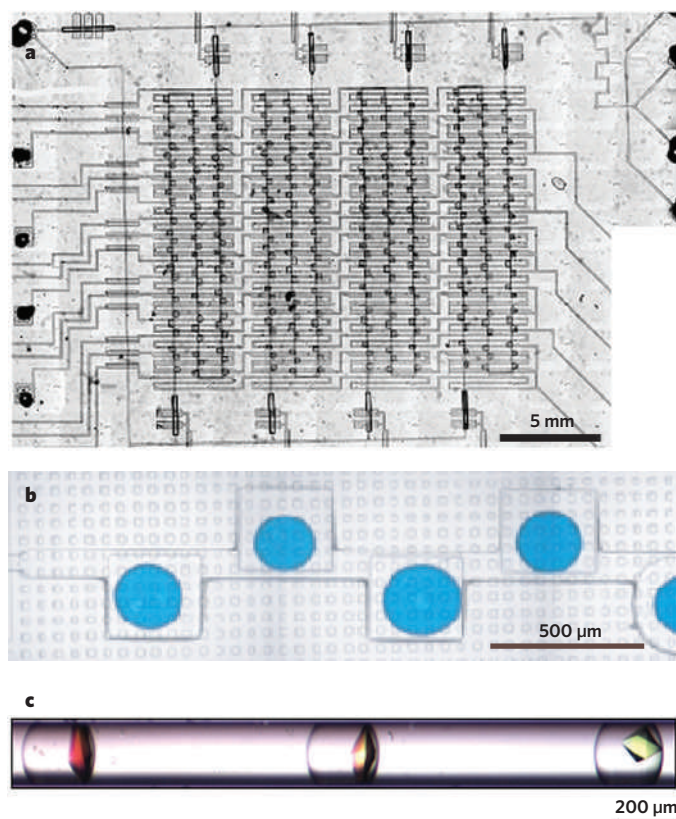


Figure 2 | Efficient screening for optimal protein crystallization conditions.

Microfluidic devices are well suited for screening conditions under which proteins crystallize, as demonstrated initially by Quake²⁰. **a**, A device in which droplets containing proteins are trapped in wells in the microchannels²¹. The droplets are then subjected to many different conditions for crystallization. **b**, An optical micrograph of dyed droplets in the wells of the device. (Images courtesy of S. Fraden and J.-u. Shim, Brandeis University, USA.) **c**, Droplets containing crystallized proteins. The droplets are produced in a microfluidic device, then collected in a glass capillary. (Image adapted, with permission, from ref. 22.)

more important than viscosity. In microsystems, with water as a fluid, the opposite is true: fluids do not mix convectively — when two fluid streams come together in a microchannel, they flow in parallel, without eddies or turbulence, and the only mixing that occurs is the result of diffusion of molecules across the interface between the fluids. Although this type of flow — known as laminar flow — requires the development of specific devices or components to accomplish mixing (when mixing is required), it has proved an advantage (and one that is characteristic of microfluidic systems) in many circumstances. The ratio of inertial to viscous forces on fluids is characterized by the Reynolds number (Re) — one of the many dimensionless parameters used in studying fluids¹⁵.

Fluids flowing in microsystems have many other interesting and useful characteristics, only some of which have been exploited. One particularly useful characteristic is electro-osmotic flow (EOF)¹⁷. When an ion-containing fluid (for example, water) is placed in a microchannel that has fixed charges on its surface (such as silicon dioxide or surface-oxidized PDMS) and an electrical potential is applied along the channel, the fluid moves as a plug, rather than with the parabolic-flow profile observed when pumping is accomplished by applying pressure to the fluid. EOF minimizes the broadening of plugs of sample that occurs with many pressure-driven systems, and allows very high resolution separations of ionic species. It is a key contributor to electrophoretic separations of DNA in microchannels¹⁸. A second potentially useful characteristic is the ability of nanofluidics to manipulate water in channels whose dimensions are similar to those of the Debye layer¹⁹: we do not, in fact, understand the characteristics of fluids at those scales, and nanofluidic systems offer windows into new phenomena in fluid physics.

Current applications

There are now enough methods of fabrication, and a sufficient range of components, to make it possible to begin to apply microfluidic systems to the resolution of problems (rather than simply to the demonstration of principles). The most highly developed of their applications is probably their use to screen conditions (such as pH, ionic strength and composition, cosolvents, and concentration) for protein crystallization^{20–22} (Fig. 2); these procedures offer the potential to screen large numbers of conditions, to separate nucleation and growth of crystals, and to minimize the damage to crystals by handling once they have formed. Some of this technology is now commercially available. Other applications for which there are laboratory demonstrations include separations coupled to mass spectroscopy²³, high-throughput screening in drug development^{24,25}, bioanalyses²⁶, examination and manipulation of samples consisting of a single cell^{27,28} or a single molecule^{29,30}, and synthesis of ¹⁸F-labelled organic compounds for positron emission tomography (PET)³¹. The area of single-molecule studies is discussed in this issue by Craighead (see page 387).

The manipulation of multiphase flows is another strength of microfluidic systems. They enable the generation and manipulation of monodisperse bubbles^{32,33} (Fig. 3) or droplets^{34–37} of a dispersed gas or liquid phase in a continuous liquid stream; these dispersions suggest new routes to the

production of polymer particles, emulsions and foams³⁸. Droplets can also serve as compartments in which to study fast organic reactions. Fluids in microchannels form the basis of new optical systems: a range of systems — from waveguides comprising a liquid with a high index of refraction flowing laminarily between two streams of low-index ‘cladding’, to applications of fluids in lenses and Bragg mirrors — are based on microfluidics^{39–44}. Psaltis, Yang and Quake paint a detailed picture of this new field, and of some of its potentials, in this issue (see page 381).

Cell biology is an area of research into which microfluidic systems bring a new capability. Jensen *et al.* (see page 403) describe types of system that seem certain to become useful new tools for cell biologists, as well as capabilities that are still needed. Eukaryotic cells, when attached and spread, have linear dimensions of 10–100 µm; these dimensions are well suited for current microfluidic devices, and PDMS — with its excellent optical transparency, low toxicity and high permeability to dioxygen and carbon dioxide — is a material that is probably uniquely suitable as a medium for the fabrication of microchambers in which to grow and observe cells^{45–48} (Fig. 4). PDMS microfluidic systems have applications in the extensive study of many areas of cell biology, including the cytoskeleton⁴⁹, the forces exerted by cells on the substrate to which they are attached⁵⁰, the contents of cells (down to the single-cell level)^{27,51}, separations of motile and non-motile cells (for example, sperm)⁵², and embryos^{53–55}.

Chemical synthesis (especially in organic and medicinal chemistry) — an area in which microfluidic systems would seem to fit naturally — has been slow to adopt microfluidic structures as a strategy for the development of new capabilities. (Some of the characteristics of chemical reactions in microsystems are discussed in this issue by deMello, page 394.) Two factors contribute to this slow adoption. First, the flexibility of conventional apparatus has, so far, not been equalled in microfluidic systems. Second, PDMS — the material most commonly used in academic studies of microfluidics — dissolves in, or is swelled by, many common organic solvents⁵⁶. The use of silicon, glass or steel^{57–59}, or perhaps polymers other than PDMS⁶⁰, may both solve this problem and allow reactions at high temperatures and pressures, but the fabrication of devices with any of these materials is more difficult than with PDMS. Pumping and valving in rigid materials such as steel must be accomplished using entirely different strategies from those used in PDMS.

The development of practical microanalytical systems^{61–64} — especially those for bioanalysis — continues rapidly, although, given its early focus, this area has been slower than expected to reach widespread routine use. Part of the problem is that there is limited technology in two parts of the cycle of analysis: sample preparation and detection. Biological samples — particularly clinical samples (such as blood or faeces), or those obtained by environmental sampling (such as soil) — are often dilute or complicated. Before these samples can be analysed by microfluidic devices, they must be converted to a form that is compatible with the intended analysis, and then introduced into the analytical device. The procedures required to complete these tasks are surprisingly sample-dependent, and not necessarily ‘micro’ in scale. After a sample has been prepared, introduced into the analytical device and processed, it must then be detected. This detection is still commonly accomplished by a microscope located off-chip. Having the microfluidic chip as just a small part of a system in which sample introduction and detection are much more complicated than the chip’s operation may be appropriate in some circumstances, but does detract from the potential advantages of microfluidic devices. Other standard problems, such as pumping, valving and on-chip reagent storage, also require better solutions than those available so far.

The future

What requirements must be fulfilled for microfluidics to become a major new technology? Will it live up to the hopes experienced at its conception? As a field, the problems it faces are those faced by most fields as they develop. The fact that microfluidics has not yet lived up to its early advertising is not a surprise, and the reasons for the rate at which it has developed are both characteristic of new technologies, and suggestive of areas in which to focus work in the future.

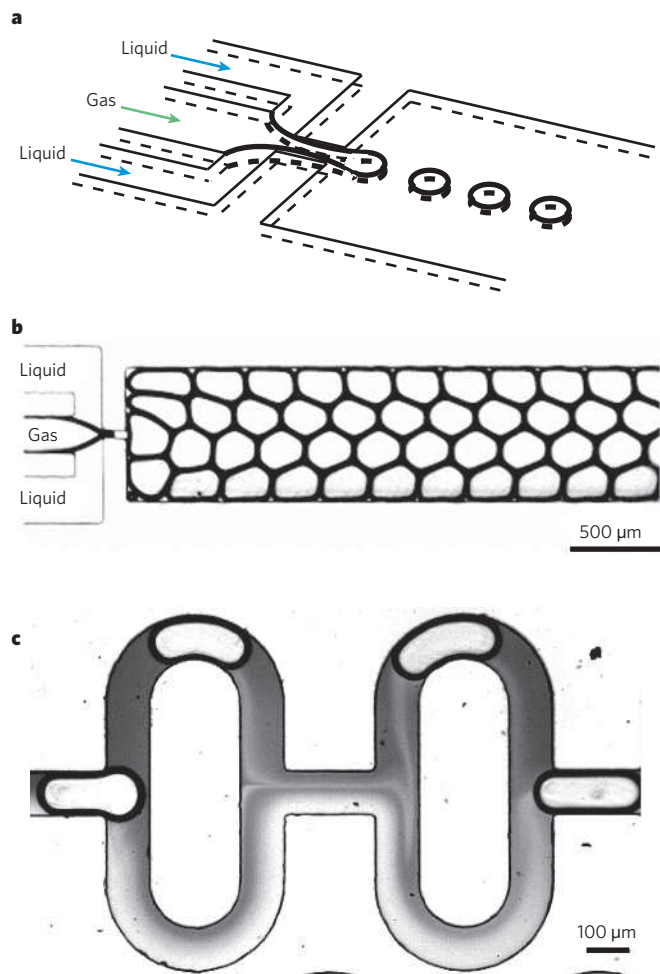


Figure 3 | Creating and using bubbles in microfluidic devices. In a microfluidic ‘flow-focusing’ device, streams of liquid pinch off a gaseous thread to produce bubbles that are remarkably monodisperse³³. The flow rate of the liquid and the pressure applied to the gas control the size of the bubbles, and the frequency with which they form. **a**, A schematic diagram of a flow-focusing system. **b**, An optical micrograph of the production of a foam comprising monodisperse bubbles. **c**, An optical micrograph of bubbles enhancing the mixing of an aqueous solution of ink (black) and an aqueous stream containing a surfactant (white).

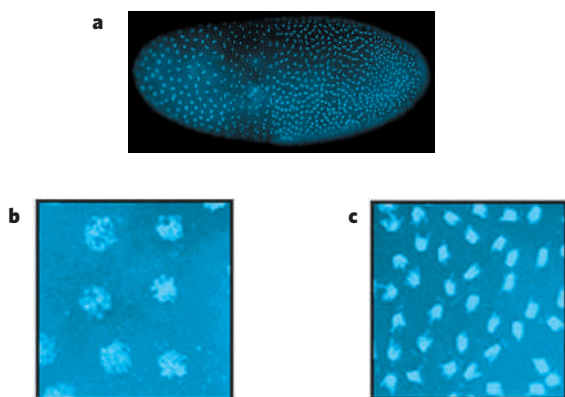


Figure 4 | A new platform for cellular and developmental biology. Laminar flow in microchannels, together with the biocompatibility of PDMS, enable new methods of studying cellular and developmental biology. One system examines the effect of temperature on the development of a fruitfly embryo⁵⁵. The embryo (the large oval in **a**) is immobilized in the middle of a microchannel. Aqueous streams of two different temperatures flow over the halves of the embryo (**b** shows the cold half, **c** the warm half); the differences in the embryo are reflected in the density of cells (marked by the light-blue nuclei). The embryo is ~500 μm wide.

General issues in introducing new technologies

The original hope for microfluidics, and that which still motivates many of us working in the field, is that it will be a practical technology — one widely used in a number of different types of application. I am confident that, ultimately, it will be, but there are several problems that it — in common with other new technologies — must first solve. Above all, it must become successful commercially, rather than remain a field based on proof-of-concept demonstrations and academic papers. The impact of microfluidic systems — as with other tools such as lasers, NMR (nuclear magnetic resonance) spectroscopy and scanning probe microscopes — will only become apparent when everyone is using them. Microfluidics must be able to solve problems for users who are not experts in fluid physics or nanolithography, such as clinicians, cell biologists, police officers or public health officials. For these applications, corporations must take on the task of making appropriate systems widely and inexpensively available.

As with all technology in transition from university laboratories to industry, the question of ‘Who owns what?’ — the problem of intellectual property — is one that must be resolved. For technology with very high value, such as biopharmaceuticals and information-processing systems, issues of intellectual property can usually be resolved by compromising on royalties, up-front payments or equity. However, some of the most interesting applications of microfluidics are those that would demand large volumes but low prices — for example, in public health monitoring, environmental monitoring, and for use in the medical systems of developing economies. In these areas, the historical differences in the valuations placed by universities and industry on university-based technology can become a serious issue: if the university places the value of an invention too high, it is simply not worthwhile to develop a commercial technology from it.

There is also the issue of the so-called ‘first-user premium’. In the introduction of a new technology, the first commercial user of that technology pays a disproportionate share of the costs of its development, and accepts a disproportionate share of the risk for that development. If the application of such a development is a very appealing — if it is of potentially high value (the ‘killer application’, or ‘killer app’) — these costs and risks are more acceptable. The high-value killer app for microfluidics has not yet emerged, although markets in research biology are certainly developing.

High-value applications

There are, in principle, high-value applications for microfluidic systems, although developing these applications requires innovations in both

microfluidics and in biomedicine; doing two things at once is always difficult. The development of new types of bioassay for monitoring patient response to therapy is one such application; development of assays for home testing, or for use in doctors’ offices at early stages of disease (early detection of ‘biomarkers’), is a second. Both are plausible developments in biomedicine, but will require both an understanding of biomarkers of disease and microfluidic systems that are highly developed. In the future, it is certainly possible that healthcare will move from treating to anticipating disease. Widespread, sensitive, frequent screening or testing will be a necessity for such anticipatory healthcare, and microfluidic systems are the most plausible technology for such testing.

Tools for the pharmaceutical industry

The pharmaceutical industry is technically sophisticated, and capable, in principle, of adopting sophisticated new technologies. The industry is also suffering from a crisis in productivity, and desperately needs new tools to guide the development of new drugs — especially to help predict the behaviour of potential new drugs in humans from performance in animals and cells. Some analytical applications of microfluidic systems in the production and use of biopharmaceuticals seem straightforward (for example, analytical systems to monitor and optimize the production of protein drugs such as therapeutic antibodies); others (such as assays based on primary human cells that could predict performance in human clinical trials) are technically more complicated, but also feasible, at least in some instances. In either case, the assays must package microsystems (almost certainly microfluidic systems) in a highly reproducible and easily manipulated format that could be used routinely by technicians.

Research

The introduction and development of new technologies is often facilitated by large, relatively cost-insensitive uses in research: equipment for processing and analysing DNA and RNA are recent examples. The development of new microfluidic tools for genomics, proteomics and metabolomics is proceeding rapidly in research laboratories, and will provide a stimulus for large-scale production.

Large-volume microanalytical tools

Among the most interesting and important of the potential applications of microfluidic systems are those toward which it was originally targeted: biomedical and related applications requiring small amounts of sample, routine operation by untrained personnel, and low cost (Fig. 5).

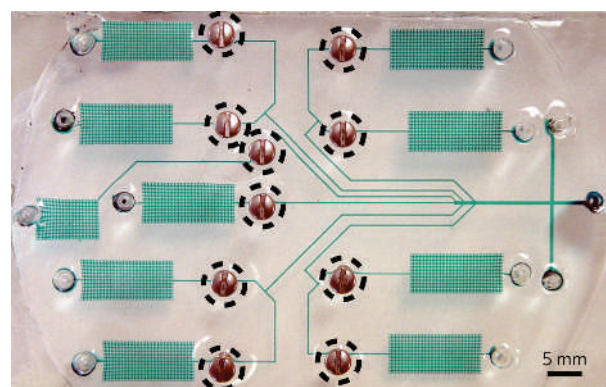


Figure 5 | A simple, inexpensive microfluidic diagnostic device.

Components of microfluidic devices can be designed to be inexpensive and easy to operate. This device⁷ performs sandwich immunoassays — tests that are used widely in medicine and biological research. The screws in this system (marked with dashed circles) act as simple, manually operated valves. Green-dyed water marks the channels. Low-cost, portable, easy-to-operate microfluidic devices such as this one may find applications in resource-poor environments. (Image adapted, with permission, from ref. 7.)

There is an appealing commonality in a number of these applications, and the volumes of appropriate analyses could, in principle, be very large (hundreds of millions of tests per year). This group of applications would include healthcare delivery and monitoring in developing economies, home healthcare and use in doctors' offices in developed economies, uses in homeland security and counterterrorism, use by first responders (police, paramedics and fire departments), applications in veterinary medicine, and incorporation into environmental and food-safety monitoring. Diagnostic systems for developing economies will require low-cost, adaptable microfluidic technology for its success; this rapidly developing field is described in this issue by Yager *et al.* (see page 412).

These applications suffer from the problem of chicken and egg: the volume of use will only be large if the cost of the analysis is low and the state of development of the assay is high; and the cost will only be low if the volume is large.

New science and technology

The development of microfluidics has just begun. A number of factors suggest that there are many early-stage applications of microsystems containing fluids, including the exploration of fluidic optics and cells, the development of new types of organic synthesis in small-channel systems, the continuing development of technologies based on large arrays of detectors and on high-throughput screening, the fabrication of microrobotic systems using hydraulic systems based on microfluidics, other fluidic versions of MEMS, and work on biomimetic systems with microfluidic components. The extension of microfluidic systems into nanofluidics — in which the dimensions of the channels and the thickness of the layer of structured fluid at the walls of the device become comparable — will make possible the exploration of the properties of near-surface water, and of ion and electrolyte transport at this interface. The biocompatibility of PDMS suggests that it might ultimately be possible to embed microfluidic devices *in vivo* for certain types of biomedically relevant analysis. Single-cell and single-molecule analysis require technologies that can work with small volumes of sample, which might allow the testing of fundamental assumptions of cell biology and molecular chemistry and biology.

Design and manufacturing systems for microfluidic devices

An important aspect of the commercial development of microfluidics — crucial to many of these applications — is the development of the technology for manufacturing microfluidic devices. Ultimately, there will probably be several such technologies, but in the early stages the definition of a single set of materials and processes needed to convert laboratory demonstrations into working commercial devices is an important step. Should devices be developed in Mylar, or PDMS, or polycarbonate? What will be the specifications for user interfaces? How important will very-large-volume technologies, such as roll-to-roll processing, be? And what about technologies for sealing and packaging?

Conclusion

So, what next for microfluidics? It is both a science and a technology. It offers great — perhaps even revolutionary — new capabilities for the future. It is also in its infancy, and a great deal of work needs to be done before it can be claimed to be more than an active field of academic research. However, the fundamentals of the field are very strong: much of the world's technology requires the manipulation of fluids, and extending those manipulations to small volumes, with precise dynamic control over concentrations, while discovering and exploiting new phenomena occurring in fluids at the microscale level, must, ultimately, be very important. ■

1. Manz, A. *et al.* Planar chips technology for miniaturization and integration of separation techniques into monitoring systems — capillary electrophoresis on a chip. *J. Chromatog.* **593**, 253–258 (1992).
2. Ng, J. M. K., Gitlin, I., Stroock, A. D. & Whitesides, G. M. Components for integrated poly(dimethylsiloxane) microfluidic systems. *Electrophoresis* **23**, 3461–3473 (2002).
3. Whitesides, G. M. & Stroock, A. D. Flexible methods for microfluidics. *Phys. Today* **54**, 42–48 (2001).

4. Mijatovic, D., Eijkel, J. C. T. & van den Berg, A. Technologies for nanofluidic systems: top-down vs. bottom-up — a review. *Lab Chip* **5**, 492–500 (2005).
5. Czaplewski, D. A., Kameoka, J., Mathers, R., Coates, G. W. & Craighead, H. G. Nanofluidic channels with elliptical cross sections formed using a nonlithographic process. *Appl. Phys. Lett.* **83**, 4836–4838 (2003).
6. Hong, J. W. & Quake, S. R. Integrated nanoliter systems. *Nature Biotechnol.* **21**, 1179–1183 (2003).
7. Weibel, D. B. *et al.* Torque-actuated valves for microfluidics. *Anal. Chem.* **77**, 4726–4733 (2005).
8. Nguyen, N. T. & Wu, Z. G. Micromixers — a review. *J. Micromech. Microeng.* **15**, R1–R16 (2005).
9. Gunther, A., Huh, J., Thalman, M., Schmidt, M. A. & Jensen, K. F. Micromixing of miscible liquids in segmented gas-liquid flow. *Langmuir* **21**, 1547–1555 (2005).
10. Garstecki, P., Fischbach, M. A. & Whitesides, G. M. Design for mixing using bubbles in branched microfluidic channels. *Appl. Phys. Lett.* **86**, 244108 (2005).
11. Laser, D. J. & Santiago, J. G. A review of micropumps. *J. Micromech. Microeng.* **14**, R35–R64 (2004).
12. McDonald, J. C. *et al.* Fabrication of microfluidic systems in poly(dimethylsiloxane). *Electrophoresis* **21**, 27–40 (2000).
13. Thorsen, T., Maerkl, S. J. & Quake, S. R. Microfluidic large-scale integration. *Science* **298**, 580–584 (2002).
14. Stone, H. A., Stroock, A. D. & Ajdari, A. Engineering flows in small devices: microfluidics toward a lab-on-a-chip. *Annu. Rev. Fluid Mech.* **36**, 381–411 (2004).
15. Squires, T. M. & Quake, S. R. Microfluidics: fluid physics at the nanoliter scale. *Rev. Mod. Phys.* **77**, 977–1026 (2005).
16. Beebe, D. J., Mensing, G. A. & Walker, G. M. Physics and applications of microfluidics in biology. *Annu. Rev. Biomed. Eng.* **4**, 261–286 (2002).
17. Santiago, J. G. Electroosmotic flows in microchannels with finite inertial and pressure forces. *Anal. Chem.* **73**, 2353–2365 (2001).
18. Wainright, A., Nguyen, U. T., Bjornson, T. & Boone, T. D. Preconcentration and separation of double-stranded DNA fragments by electrophoresis in plastic microfluidic devices. *Electrophoresis* **24**, 3784–3792 (2003).
19. Karnik, R., Castelino, K. & Majumdar, A. Field-effect control of protein transport in a nanofluidic transistor circuit. *Appl. Phys. Lett.* **88**, 123114 (2006).
20. Hansen, C. L., Skordalakes, E., Berger, J. M. & Quake, S. R. A robust and scalable microfluidic metering method that allows protein crystal growth by free interface diffusion. *Proc. Natl Acad. Sci. USA* **99**, 16531–16536 (2002).
21. Shim, J.-u., Cristobal, G., Link, D. R., Thorsen, T. & Fraden, S. Using microfluidics to decouple nucleation and growth of protein crystals. *J. Amer. Chem. Soc.* (submitted).
22. Zheng, B., Tice, J. D., Roach, L. S. & Ismagilov, R. F. A droplet-based, composite PDMS/glass capillary microfluidic system for evaluating protein crystallization conditions by microbatch and vapor-diffusion methods with on-chip X-ray diffraction. *Angew. Chem. Int. Ed.* **43**, 2508–2511 (2004).
23. Ramsey, R. S. & Ramsey, J. M. Generating electrospray from microchip devices using electroosmotic pumping. *Anal. Chem.* **69**, 1174–1178 (1997).
24. Dittrich, P. S. & Manz, A. Lab-on-a-chip: microfluidics in drug discovery. *Nature Rev. Drug Discov.* **5**, 210–218 (2006).
25. Pihl, J., Karlsson, M. & Chiu, D. T. Microfluidic technologies in drug discovery. *Drug Discov. Today* **10**, 1377–1383 (2005).
26. Sia, S. K. & Whitesides, G. M. Microfluidic devices fabricated in poly(dimethylsiloxane) for biological studies. *Electrophoresis* **24**, 3563–3576 (2003).
27. Wheeler, A. R. *et al.* Microfluidic device for single-cell analysis. *Anal. Chem.* **75**, 3581–3586 (2003).
28. Werdich, A. A. *et al.* A microfluidic device to confine a single cardiac myocyte in a sub-nanoliter volume on planar microelectrodes for extracellular potential recordings. *Lab Chip* **4**, 357–362 (2004).
29. Dittrich, P. S. & Manz, A. Single-molecule fluorescence detection in microfluidic channels — the Holy Grail in μ TAS? *Anal. Bioanal. Chem.* **382**, 1771–1782 (2005).
30. Stavits, S. M., Edel, J. B., Samiee, K. T. & Craighead, H. G. Single molecule studies of quantum dot conjugates in a submicrometer fluidic channel. *Lab Chip* **5**, 337–343 (2005).
31. Lee, C. C. *et al.* Multistep synthesis of a radiolabeled imaging probe using integrated microfluidics. *Science* **310**, 1793–1796 (2005).
32. Ganan-Calvo, A. M. & Gordillo, J. M. Perfectly monodisperse microbubbling by capillary flow focusing. *Phys. Rev. Lett.* **87**, 274501 (2001).
33. Garstecki, P. *et al.* Formation of monodisperse bubbles in a microfluidic flow-focusing device. *Appl. Phys. Lett.* **85**, 2649–2651 (2004).
34. Thorsen, T., Roberts, R. W., Arnold, F. H. & Quake, S. R. Dynamic pattern formation in a vesicle-generating microfluidic device. *Phys. Rev. Lett.* **86**, 4163–4166 (2001).
35. Link, D. R., Anna, S. L., Weitz, D. A. & Stone, H. A. Geometrically mediated breakup of drops in microfluidic devices. *Phys. Rev. Lett.* **92**, 054503 (2004).
36. Anna, S. L., Bontoux, N. & Stone, H. A. Formation of dispersions using “flow focusing” in microchannels. *Appl. Phys. Lett.* **82**, 364–366 (2003).
37. Tan, Y. C., Fisher, J. S., Lee, A. I., Cristini, V. & Lee, A. P. Design of microfluidic channel geometries for the control of droplet volume, chemical concentration, and sorting. *Lab Chip* **4**, 292–298 (2004).
38. Xu, S. Q. *et al.* Generation of monodisperse particles by using microfluidics: control over size, shape, and composition. *Angew. Chem. Int. Ed.* **44**, 724–728 (2005).
39. Wolfe, D. B. *et al.* Dynamic control of liquid-core/liquid-cladding optical waveguides. *Proc. Natl Acad. Sci. USA* **101**, 12434–12438 (2004).
40. Kerbage, C. & Eggleton, B. J. Tunable microfluidic optical fiber gratings. *Appl. Phys. Lett.* **82**, 1338–1340 (2003).
41. Datta, A. *et al.* Microfabrication and characterization of Teflon AF-coated liquid core waveguide channels in silicon. *IEEE Sens. J.* **3**, 788–795 (2003).
42. Balslev, S. & Kristensen, A. Microfluidic single-mode laser using high-order Bragg grating and antiguiding segments. *Opt. Express* **13**, 344–351 (2005).
43. Campbell, K. *et al.* A microfluidic 2x2 optical switch. *Appl. Phys. Lett.* **85**, 6119–6121 (2004).

44. Vezenov, D. V. *et al.* A low-threshold, high-efficiency microfluidic waveguide laser. *J. Am. Chem. Soc.* **127**, 8952–8953 (2005).
45. Hung, P. J., Lee, P. J., Sabounchi, P., Lin, R. & Lee, L. P. Continuous perfusion microfluidic cell culture array for high-throughput cell-based assays. *Biotechnol. Bioeng.* **89**, 1–8 (2005).
46. Chung, B. G. *et al.* Human neural stem cell growth and differentiation in a gradient-generating microfluidic device. *Lab Chip* **5**, 401–406 (2005).
47. Taylor, A. M. *et al.* Microfluidic multicompartiment device for neuroscience research. *Langmuir* **19**, 1551–1556 (2003).
48. Walker, G. M. *et al.* Effects of flow and diffusion on chemotaxis studies in a microfabricated gradient generator. *Lab Chip* **5**, 611–618 (2005).
49. Takayama, S. *et al.* Selective chemical treatment of cellular microdomains using multiple laminar streams. *Chem. Biol.* **10**, 123–130 (2003).
50. Lu, H. *et al.* Microfluidic shear devices for quantitative analysis of cell adhesion. *Anal. Chem.* **76**, 5257–5264 (2004).
51. McClain, M. A. *et al.* Microfluidic devices for the high-throughput chemical analysis of cells. *Anal. Chem.* **75**, 5646–5655 (2003).
52. Cho, B. S. *et al.* Passively driven integrated microfluidic system for separation of motile sperm. *Anal. Chem.* **75**, 1671–1675 (2003).
53. Walters, E. M., Clark, S. G., Beebe, D. J. & Wheeler, M. B. Mammalian embryo culture in a microfluidic device. *Methods Mol. Biol.* **254**, 375–382 (2004).
54. Glasgow, I. K. *et al.* Handling individual mammalian embryos using microfluidics. *IEEE Trans. Biomed. Eng.* **48**, 570–578 (2001).
55. Lucchetta, E. M., Lee, J. H., Fu, L. A., Patel, N. H. & Ismagilov, R. F. Dynamics of *Drosophila* embryonic patterning network perturbed in space and time using microfluidics. *Nature* **434**, 1134–1138 (2005).
56. Lee, J. N., Park, C. & Whitesides, G. M. Solvent compatibility of poly(dimethylsiloxane)-based microfluidic devices. *Anal. Chem.* **75**, 6544–6554 (2003).
57. Jensen, K. F. Silicon-based microchemical systems: characteristics and applications. *MRS Bull.* **31**, 101–107 (2006).
58. Lowe, H. & Ehrfeld, W. State-of-the-art in microreaction technology: concepts, manufacturing and applications. *Electrochim. Acta* **44**, 3679–3689 (1999).
59. Snyder, D. A. *et al.* Modular microreaction systems for homogeneously and heterogeneously catalyzed chemical synthesis. *Helv. Chim. Acta* **88**, 1–9 (2005).
60. Rolland, J. P., Van Dam, R. M., Schorzman, D. A., Quake, S. R. & DeSimone, J. M. Solvent-resistant photocurable “liquid Teflon” for microfluidic device fabrication. *J. Am. Chem. Soc.* **126**, 2322–2323 (2004).
61. Auroux, P. A., Koc, Y., deMello, A., Manz, A. & Day, P. J. R. Miniaturised nucleic acid analysis. *Lab Chip* **4**, 534–546 (2004).
62. Breslauer, D. N., Lee, P. J. & Lee, L. P. Microfluidics-based systems biology. *Mol. Biosys.* **2**, 97–112 (2006).
63. Huh, D., Gu, W., Kamotani, Y., Grotberg, J. B. & Takayama, S. Microfluidics for flow cytometric analysis of cells and particles. *Physiol. Meas.* **26**, R73–R98 (2005).
64. Suh, R., Takayama, S. & Smith, G. D. Microfluidic applications for andrology. *J. Androl.* **26**, 664–670 (2005).
65. Balagaddé, F. K., You, L., Hansen, C. L., Arnold, F. H. & Quake, S. R. Long-term monitoring of bacteria undergoing programmed population control in a microchemostat. *Science* **309**, 137–140 (2005).

Acknowledgements I thank M. Fuerstman for extensive help in preparing this manuscript. This work was supported by a grant from the National Institutes of Health.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence should be addressed to the author (gwhitesides@gmwgroup.harvard.edu).

Scaling and the design of miniaturized chemical-analysis systems

Dirk Janasek¹, Joachim Franzke¹ & Andreas Manz¹

Micrometre-scale analytical devices are more attractive than their macroscale counterparts for various reasons. For example, they use smaller volumes of reagents and are therefore cheaper, quicker and less hazardous to use, and more environmentally appealing. Scaling laws compare the relative performance of a system as the dimensions of the system change, and can predict the operational success of miniaturized chemical separation, reaction and detection devices before they are fabricated. Some devices designed using basic principles of scaling are now commercially available, and opportunities for miniaturizing new and challenging analytical systems continue to arise.

In chemical engineering, problems frequently arise in scaling up chemical processes. Research is normally conducted in glassware on the millilitre scale, whereas cubic-metre capacities are required for production. The main scale-up problems are associated with heat and mass transport, and can result in increased formation of by-products and lower yields. In the worst cases, shortcomings can lead to runaway, in which the rate of heat generation exceeds the rate of cooling available, and other hazardous situations.

Some of the scaling laws described here were developed between the 1880s and the 1930s in the growing field of engineering. The aim was to provide a framework for engineers to establish how material would behave on different length scales, allowing them to optimize output and minimize the risk of runaway and other hazards. In the pre-computer era, these laws proved to be simple and useful tools, particularly when the engineering mathematics required to model a chemical process became complicated.

During the past 20 years, microfluidics, micrometre-scale total analysis systems (μ TAS) or so-called 'lab-on-a-chip' devices have revived interest in these scaling laws and dimensionless groups for downscaling purposes¹. Such devices have a range of practical benefits (see page 368).

Here, we aim to review some of the important principles that contribute to the design of novel μ TAS and to propose future research activity, to inform the reader, who might ultimately use such devices for research, and even to inspire the reader to take on the challenge of designing new μ TAS.

For simplicity, we use three miniaturized devices as the main examples for discussion: an open-tubular chromatographic system that is used to separate molecules from mixtures (Fig. 1); a microwell plate as an example of a device in which chemical interactions must be optimized; and a gas-phase detection device that uses a glow-discharge plasma (Fig. 2).

The scaling laws considered here are generally not valid on the nanometre scale, so 200 years' experience in chemistry cannot be applied in this case. We are not yet convinced that practical applications of microfluidics are feasible; however, there seems to be tremendous potential for basic research on the subject. We consider the absolute limits of the principles described here for the scaling down of chemical processes.

Separation

The measurement of a specific compound in a complex sample matrix can be carried out using selective (bio)chemical sensors or non-specific detectors applied after the separation of the sample mixture into

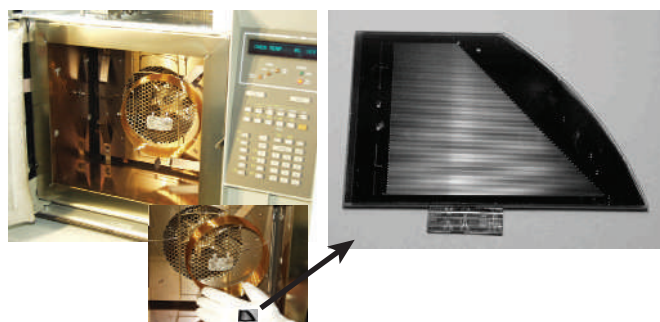


Figure 1 | Examples of the miniaturization of separation techniques. The figure compares the size of a commercial gas chromatograph and column with that of a microscaled column on a chip. (Photo of microscaled column courtesy of J. Müller, Institut für Mikrosystemtechnik, Technische Universität Hamburg-Harburg, Germany.)

distinct zones of the single analytes. The most widely used separation principles are chromatography (based on the different distributions of the compounds within a complex mixture between two phases) and electrophoresis (based on the differential movement of charged components in an electric field).

Some of the initial theoretical work on scaling down devices was in the field of chromatography and was published as early as the 1950s — for example, by Golay on open-tubular gas chromatography² and van Deemter *et al.* on packed-column liquid chromatography³. These papers led to the development of commercial gas chromatographs using capillaries with micrometre diameters, and liquid chromatographs with micrometre-scale particles for the stationary phase.

The first commercial microfluidic-based platform for the analysis of DNA, RNA, proteins and cells — known as the 2100 Electrophoresis Bioanalyzer — was launched in 1999 by Agilent. Based on the LabChip technology from Caliper LifeSciences, it provides good-quality data quickly and accurately and is an alternative to labour-intensive gel electrophoresis. Using one platform, the integrity and purity of RNA can be determined in less than 30 min, 16 samples of RT-PCR (polymerase chain reaction with reverse transcription) products can be quantified in the same time range, the transfection efficiency of cloning experiments

¹ISAS — Institute for Analytical Sciences, Bunsen-Kirchhoff-Str. 11, D-44139 Dortmund, Germany.

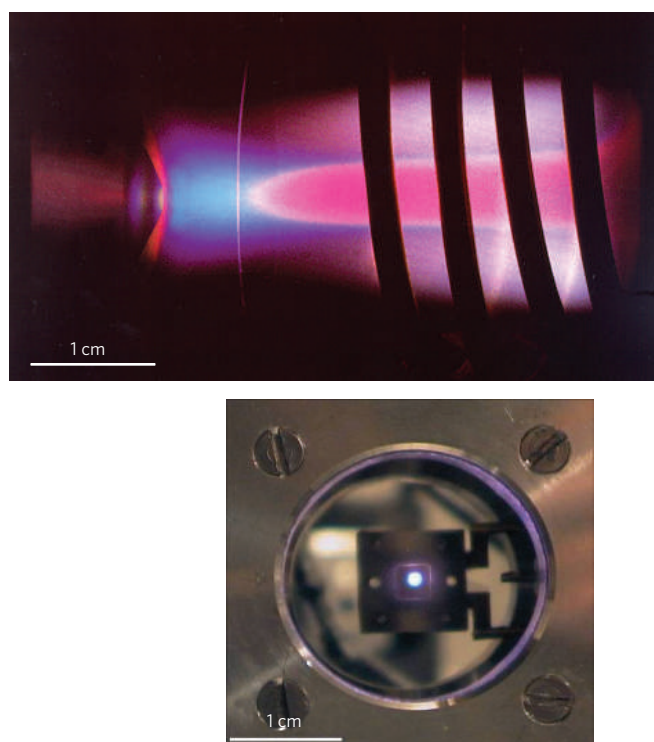


Figure 2 | Plasma detectors, an example of miniaturization in detection. Photographs of an inductively coupled plasma with power consumption of 1 kW and a microhollow cathode plasma of less than 1 W. The whole of the microhollow cathode has a diameter of 100 μm .

can be monitored, and the size and concentration of expressed proteins can be analysed. The success and reliability of the 2100 Bioanalyzer can be estimated from the number of publications in which the system has been used. Since its launch, the number has increased exponentially and reached 150 per month in 2005 (G. P. Rozing, personal communication). The device demonstrates the success of the application of scaling laws in chemical separations.

Two approaches have proved particularly useful in the design of miniature separation devices, and these are discussed below.

Dimensionless-parameters approach

'Dimensionless parameters' define units such as volume, column length, linear flow rate, retention time and pressure drop in terms of quantities that can be assumed to be constant over the entire system. For decades, engineers have used dimensionless parameters to correlate experimental results when large numbers of significant variables are involved.

In the case of chromatography using an open-tubular column (in which the columns are not packed, but instead the walls of the column support the stationary phase and a hollow core allows the mobile phase

to move freely), the constant quantities include inner diameter, mobile-phase viscosity, average diffusion coefficient of a sample in the mobile phase, and Poiseuille number for a circular cross-section. The dimensionless-parameters approach can be applied as long as the contribution of other physical phenomena not considered in the parameter can be neglected.

Nonlinear behaviour can be usefully described with the aid of dimensionless parameters⁴ such as the Peclet number, ν , which is the ratio of axial bulk flow to diffusion mass transport (effectively describing the relative importance of convective or diffusive transport, or change in flow rate), the Fourier number, τ , which describes the average number of times a molecule contacts the wall of the capillary (elution time, or mixing efficiency), and the Bodenstein number, Π , which characterizes the backmixing within a system (or pressure drop). These parameters make it possible to extrapolate results obtained for one system to other similar systems through multiplication by constant factors, as shown in Table 1. So, even before starting an experiment, we can establish whether or not the device will function. For instance, miniaturization to channel sizes of 1.5 μm results in a pressure drop of 3,900 bar, which is difficult if not impossible to provide. An example of how dimensionless variables can be applied to capillary separation systems is provided by Knox and Gilbert⁵, who derived sets of optimal conditions for capillary liquid chromatography using reduced parameters.

The Reynolds number, Re , is a particularly useful dimensionless parameter. It represents the ratio of inertial forces ($\nu\rho$, where ν is the average velocity and ρ the fluid density) to viscous forces (η/d , where η is the dynamic viscosity and d the characteristic length, such as the diameter of the capillary); it characterizes the fluid flow as being laminar or turbulent⁶. Once an experiment indicates that a transition will occur, say at $Re = 2,000$, the outcome of any other experiment and the behaviour of the flow can be predicted. Because the Reynolds number is proportional to the characteristic length d , for miniaturized systems, a small value will be obtained, indicating laminar flow. Such parameters are useful for controlling separations and for mixing to optimize the output of chemical reactions (see below).

Some behaviour on the nanometre scale can be estimated by applying these particular scaling laws, and the results used as a means of testing whether answers derived by more sophisticated methods are reasonable. Perhaps surprisingly, many scaling laws for the most mechanical systems are quite accurate on the nanometre scale, but in electromagnetic systems many scaling laws fail dramatically. Scaling laws for thermal systems have variable accuracy⁷.

Scaling relationships based on classical continuum models ultimately break down as a consequence of atomic-scale structure, mean-free-path effects and quantum-mechanical effects.

Similarity approach

Useful information about the behaviour of simple miniaturized flow systems can be attained by considering the proportionalities within a system. The parameters of interest, such as flow rate, pressure drop or electric field, are viewed as a function of the variables to be miniaturized in space and time. Using this approach, the major trends

Table 1 | Examples of the dimensionless-parameters approach

Parameters	Diameter, d_c (μm)			Dimensionless parameters Relationship	Calculated values
	30	5	1.5		
Length, L (m)	13.5	2.3	0.67	Reduced column length, $\lambda = L/d_c$	450,000
Time, t (min)	122	3.13	0.28	Fourier number (reduced retention time), $\tau = tD/d_c^2$	7,500
Pressure difference, Δp (bar)	10	350	3,900	Bodenstein number (reduced pressure drop), $\Pi = p \cdot d_c^2 / \Phi \cdot \eta_m \cdot D_m$	280
Peak standard deviation in terms of length, σ_x (μm)	13,400	2,225	670	Reduced variance of elution profile in terms of length, $s_x = \sigma_x/d_c$	450
Peak standard deviation in terms of time, σ_t (ms)	6,750	188	17	Reduced variance in terms of time, $s_t = \sigma_t \cdot D_m / d_c^2$	7.5
Peak standard deviation in terms of volume, σ_v (pl)	10,200	47	1.3	Reduced variance in terms of volume, $s_v = \sigma_v / d_c^3$	350

Calculated parameter sets for an open-tubular column liquid chromatography (LC) system with one million theoretical plates at zero retention as capillary diameter changes (Peclet number, $\nu = 38$). The diffusion coefficient is assumed to be $D_m = 10^{-9} \text{ m}^2 \text{ s}^{-1}$; the viscosity is $\eta_m = 10^{-3} \text{ N s m}^{-2}$; and the Poiseuille number, Φ , has a value of 32 for a circular cross-section. An open-tubular LC system with zero retention was chosen for numerical simplicity; ν_m molecular.

a parameter undergoes during its downscaling become apparent, with no knowledge of material constants (such as viscosity or heat capacity) being necessary.

For any system, it is possible to proceed from a given point, such as an experimental result, and extrapolate to estimate the magnitude of the variables in a downscaled system. Changes in geometry only influence this estimation by a constant factor. Although this approach cannot be used to predict the feasibility of a system, it can, in principle, allow physically impossible cases to be rejected and provide an idea of the order of magnitude of relevant variables. If it is assumed that miniaturization is a simple three-dimensional downscaling process characterized by a typical length parameter, d (refs 1, 8), we can easily predict the behaviour of the relevant physical variables. The typical length d represents the scaling factor of the miniaturization. A single degree of freedom for the mechanical parameters remains: time. For simplicity, only two important cases are considered and discussed below.

If the timescale is the same for the miniaturized system as for the full-scale system, it is referred to as a time-constant system (Table 2). Relevant time variables (such as analysis time, transport time and response time) remain the same. However, linear flow rate in a tube would decrease by a factor d , volumetric flow rate by d^3 , and the Reynolds number by d^2 . By contrast, the pressure drop needed to maintain the desired flow rate would remain the same. This time-constant behaviour is important for simple transportation and for analytical techniques such as flow-injection analysis (FIA) systems. Diffusion would certainly have a more dominant role in mass flow in miniaturized systems. The main advantage in downscaling simple transport or FIA systems lies in the conservation of carrier and reagent solutions. A 10-fold decrease in size, for example, would cause a 1,000-fold decrease in carrier or reagent consumption.

The diffusion-related system (Table 2) becomes important when molecular diffusion, heat diffusion or flow characteristics dictate the separation efficiency in a given system. In such instances, the timescale is treated as a surface that is proportional to d^2 . This behaviour is in perfect agreement with standard chromatographic and electrophoretic band-broadening theory. All dimensionless parameters, including the Peclet number, Fourier number and Bodenstein number, remain constant regardless of the size of the system¹. In other words, hydrodynamic diffusion, heat diffusion and molecular diffusion effects behave in the miniaturized system exactly as in the original system.

This means that downscaling to one-tenth of the original tube diameter reduces related time variables such as analysis time and required response time for a detector to one-hundredth of their original magnitudes. The pressure drop requirements increase by a factor of 100, but the voltage requirements (for electrophoresis or electro-osmosis) remain unchanged. The main advantage of the diffusion-controlled system is that miniaturization achieves faster separations while maintaining comparable separation efficiency. An initial experiment in

this field was conducted in 1992 (ref. 9). Finally, the predictions of similarity laws (for example, the same performance of a separation in a shorter time) were experimentally validated when electrophoresis-based separations of amino acids with up to 75,000 theoretical plates could be obtained in about 15 seconds using a microchip system¹⁰. More recent work on the micrometre scale has confirmed those results for field-inversed electrophoresis¹¹, chiral separations¹² and free-flow electrophoresis (FFE)¹³.

FFE is a convenient separation technique with great potential for integration with sample preparation, detection and even chemical reactions in this micrometre-scale regime. In contrast to capillary electrophoresis (CE), in which a longitudinal separation is obtained, in FFE the time domain is converted to a spatial domain. This means that the sample solution can be fed continuously into the separation compartment, separated into its components and supplied to further actions. Applying the miniaturization to FFE devices, the aim of fast acquisition of qualitative and quantitative data can be achieved and has been demonstrated for zone electrophoretic^{14–18}, isoelectric focusing^{13,17,19} and isotachophoretic modes²⁰. We envisage that, in the future, two compounds could react in the free-flow device to form the product, which will be separated immediately and continuously from the not-yet-converted starting substances. The product can be fractionated and transferred to subsequent processes — for example, mass spectrometry/mass spectrometry detection — while starting reagents that have yet to be converted can be returned to the inlet of the free-flow device.

By downscaling further into the nanometre range, other physical phenomena have to be taken into consideration, which means that the similarity laws become invalid. As an example, an electric double layer can be considered that is formed as the negative surface charges of the capillary wall are compensated by positive ions from the buffer solution. It consists of a rigid ‘Stern’ layer in proximity to the capillary surface, and a diffuse layer extending into the bulk solution in the dimension of a few hundred nanometres. If in a nanochannel the electric double layers overlap, the streaming potential will decrease, co-ions will be excluded from the channel, and counter-ions will be enriched. Another phenomenon is the generation of capillary-induced negative pressure during two-phase flow in nanochannels: Tas *et al.*²¹ observed a peculiarly shaped meniscus of water plugs in 100-nm nanochannels, which they attributed to a downward bending of the channel capping under the influence of the tensile capillary forces. The interested reader is referred to a comprehensive overview of physical phenomena on the nanometre scale²². Articles on the promise of nanotechnology for separation devices²³ and methods of manipulating individual molecules have been published recently (see page 387).

The absolute limit of miniaturization with respect to separation techniques depends on the maximum applicable pressure (which is dependent on the stability of the used material); in the case of chromatography, on the maximum voltage that can be applied in the case of electrophoresis, and on the molecular characteristics that affect, for instance, steric behaviour, viscosity and surface tension. We emphasize that it is impossible to miniaturize further than to the level of a single molecule.

Chemical reactions

Once the requisite molecule has been separated from its complex starting matrix, manipulation of the molecules for a desired outcome (for example, synthesis or information generation) can be considered.

Miniaturization with respect to reactions first has to take into consideration the fact that a necessary concentration of molecules is required so that they have a chance of colliding and reacting. In addition, forces such as surface tension or adhesion must not prevent the collision of these few molecules.

Mixing

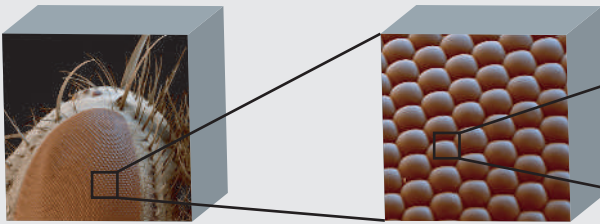
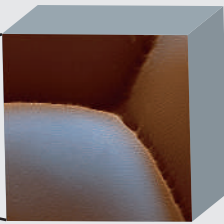
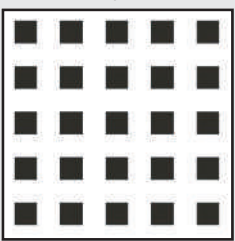
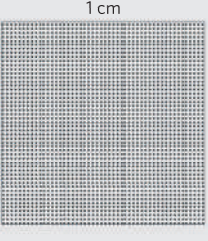
Many reactions, such as bioassays, phase-transfer reactions and multi-component reactions including PCR and the Ugi reaction (organic

Table 2 | Examples of the similarity approach

Parameter-related system	Time-constant system	Diffusion-related system
Space, x	d	d
Time, t	Constant	d^2
Linear flow rate, u	d	$1/d$
Volume flow rate, F	d^3	d
Pressure drop (laminar flow), Δp	Constant	$1/d^2$
Voltage (electro-osmotic flow), U	d^2	Constant
Electric field, U/L	d	$1/d$
Reynolds number, Re	d^2	Constant
Peclet number (reduced flow rate), ν	d^2	Constant
Fourier number (reduced elution time), τ	$1/d^2$	Constant
Bodenstein number (reduced pressure), Π	d^2	Constant
Reduced voltage, V	Constant	Constant

Proportionality factors are given for mechanical parameters in relation to a characteristic length, d .

Table 3 | Size dependent diffusion time and information density

Device characteristic	Length, d		
	1 mm	100 μm	10 μm
Volume, V	1 μL	1 nL	1 pL
			
Number of molecules at 1 μM	6×10^{11}	6×10^8	6×10^5
Diffusion time, $t_d = Dd^2$	15 min	10 s	100 ms
Number of volumes (volumes cm^2)	$5 \times 5 = 25$	$50 \times 50 = 2,500$	$500 \times 500 = 25 \times 10^4$
			
Maximum information density	1.5 per min cm^2	250 per s cm^2	2.5×10^6 per s cm^2

A number of device characteristics at three values of the typical length d . Diffusion coefficient $D = 10^{-9} \text{ m}^2 \text{ s}^{-1}$. A housefly is used to provide a pictorial representation of scale. (Reprinted, with permission, from Eye of Science.)

reactions that can be used to form libraries of low-molecular-mass drug-like compounds), are diffusion related. In Table 3, a reaction system of two reactants is described. Usually, two molecules meet by brownian motion, which is dependent on the diffusion coefficient. That meeting process, which is a mixing by diffusion, is greatly enhanced in small structures because the time a molecule needs to travel a distance d decreases as $1/d^2$. Short diffusion times in the millisecond range indicate that an efficient mixing of two solutions will be obtained when they are brought into contact on the micrometre scale. Because there is no turbulence in the micrometre regime, much effort has gone into designing devices to improve mixing capability. A mixing device based on flow lamination²⁴ and measurements of protein folding by a pH jump²⁵ used the short diffusion time on the microscale. However, the mixing process can be enhanced further by the use of chaotic advection, as demonstrated by Stroock *et al.* using herringbone mixers²⁶, and by Song *et al.* for mixing in droplets with a two-phase flow²⁷.

Multiple reactions

An important consideration in the design of systems to carry out multiple chemical analyses is the number of single, independent devices that can be arranged on a certain area — for example, on a microwell plate. This number increases with $1/d^2$. For instance, with a length of 10 μm , 250,000 devices can be arranged per square centimetre. Such a host of devices could be used in discrete applications, or together for parallel or sequential processing. An impressive example is the multistep synthesis of radiolabelled image probes²⁸. Five sequential processes can proceed with high radiochemical yield and purity on the nanogram to microgram scale, and with shorter synthesis time relative to conventional automated synthesis.

Shorter diffusion times increase the exchange of molecular information on the device and therefore the rate of information generation. When this rate is multiplied by the number of volumes present, a maximum information-density number can be defined per time and per surface area. Table 3 shows that it increases with the fourth power of the inverted distance d . It should be noted, however, that this limit cannot

be reached in all cases. Slow kinetics, fluid-handling constraints and detection requirements will all restrict information generation. High information densities will be useful in the evaluation of the millions of compounds produced by combinatorial chemistries, or in speeding up clinical DNA diagnostics.

An example is the Affymetrix GeneChip microarray. Semiconductor fabrication techniques, solid-phase and combinatorial chemistry, and molecular biology are integrated to create arrays with millions of probes for DNA–RNA hybridization experiments^{29,30}. By 2005, more than 4,200 publications reflected the extensive use of GeneChip systems in many areas, such as sequence analysis, targeted genotyping, expression quantification and regulation, as well as clinical research and molecular diagnostics (information obtained by Affymetrix, 2006).

A second example is the commercially available protein-crystallization chip launched by Fluidigm in 2003. The chip uses micrometre-scale channels based on multilayer soft lithography, and active valves for diffusive mixing of protein and crystallization reagents to overcome two major bottlenecks for protein-structure determination by X-ray crystallography: producing sufficient quantities of material and finding appropriate crystallization conditions. The TOPAZ Crystallizer chip can screen 96 crystallization conditions on four proteins in 384 parallel reactions using just 10 nL of protein per reaction; conventionally, 1 μL per experiment is needed. Here, high-throughput screening and sample-volume requirements reflect the advantages predicted by the similarity laws of scaling.

Detection

Decreasing the dimensions of reaction systems to small volumes with small amounts of analytes naturally stimulates the demand for adequate, high-sensitivity detection techniques. Besides electrochemical methods and mass spectrometry, optical techniques such as absorption, refractive-index variation, chemiluminescence and fluorescence measurements are usually used, because they are non-invasive and provide high temporal and spatial resolution for a suitable experimental set-up^{31–33}.

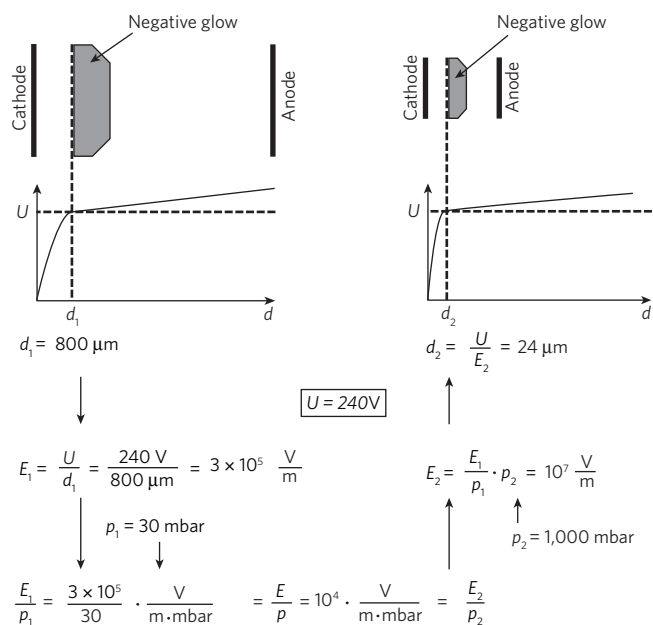


Figure 3 | Similarity relations between glow discharges on the macro- and microscales. Two geometrically similar systems are shown with cathode, anode and negative glow. Underneath, the potential of the discharge is given, depending on the distance from the cathode. The distances d_1 and d_2 are the points at which the discharge voltage change is maximal. Just as d attains values of d_1 or d_2 , the breakdown flash occurs, which is where the negative glow of the discharge is located. On the left side is the parameter d_1 , obtained with helium direct-current discharge at a pressure, p_1 , of 30 mbar. Following the arrows, the calculations show that with the application of a voltage, U , of 240 V and a pressure, p_2 , of 1,000 mbar, the distance from the cathode to the negative glow will be diminished to 24 μm . The product $p \cdot d$, the ratio E/p (E is the electric field), and all quantities that are functions of $p \cdot d$ or E/p , are the same in the two systems. The dependence of the breakdown potential on the product $p \cdot d$ was first established by de la Rue and Muller⁶⁶ in 1880. Later, Paschen concluded from an extensive study of air, CO_2 and H_2 over a range of values of $p \cdot d$ that breakthrough voltage, U_b , is a function $\psi(p \cdot d)$ of the product $p \cdot d$ only; this result is known as Paschen's law. Carr⁶⁷ confirmed its validity for a number of different gases for values of $p \cdot d$ from 0.1 Torr-cm to 15 Torr-cm. From typical Paschen⁴¹ curves, E/p is shown to diminish as $p \cdot d$ increases. Transformations of quantities in homologous regions are given in ref. 46. In considering similarity for high-frequency fields, the parameter is no longer E/p only, but also involves a time factor that must be changed by a proportionality factor, K . This parameter must contain the term f/p (f is the frequency of the field) as well as E/p .

Optical methods

Fluorescence analysis is particularly attractive because fluorophores can be excited and detected selectively. Furthermore, the excellent sensitivity of fluorescence spectroscopy is greatly enhanced by reducing the size of the detection volume, because the background signal that is generated by impurities of the sample — for example, Rayleigh stray-light and Raman scattering — scales linearly with the size of the detection volume. The fluorescence signal of a single molecule, on the other hand, is independent of the dimensions of the detection volume and remains constant^{34,35}. The resulting high signal-to-noise ratio facilitates the recognition of single fluorophores residing in the detection volume. However, with fewer than 10 molecules, the measurement would not allow a statistically covered analysis but give a digitalized output.

Assuming a molecular weight of about 50,000, analyte concentrations in the higher parts per million range are needed for detection volumes lower than a cubic micrometre. At this concentration, effects such as viscosity and light scattering have to be taken into consideration. Detection methods such as surface plasmon resonance or detections related to evanescent fields do not depend on geometrical scaling but on the wavelength of the light; however, they are subject to the same

restrictions with respect to statistical coverage. These methods will only strictly fulfil scaling criteria if the mechanical confinement becomes smaller than the light wavelength.

Mass spectrometry

Significant progress has been made in the development of miniaturized mass spectrometers in a variety of field applications ranging from assessing the composition of planetary atmospheres and monitoring air quality on manned space missions³⁶ to chemical analysis in unmanned underwater vehicles³⁷ and the environmental analysis of air^{38,39}.

Plasma discharges are used as excitation sources for molecular mass-spectrometry components and adhere to one of the oldest scaling laws we know. It is a similarity approach firstly noticed in 1915 (ref. 40). The author of this paper, Townsend⁴⁰, showed that an even earlier law⁴¹, from 1889, was a special case of a more general similarity approach that can be applied not only to breakdown in uniform fields, but also to a breakdown that depends on ionization by collision in non-uniform fields. As shown in Fig. 3, this approach can identify the smallest dimension for which a plasma discharge can be sustained under atmospheric pressure. For helium the dimension is 24 μm ; in the case of argon it is 6 μm . For analytical purposes, it makes no sense to work with higher pressures. This theorem was later extended by Holm⁴² to account for the maintenance of current between geometrically similar electrodes, and its relation to fundamental processes has been discussed by von Engel and Steenbeck⁴³. Margenau⁴⁴ showed theoretically how the principle might be extended to high-frequency alternating fields, and this has also been verified experimentally^{45,46}.

Most microplasmas developed for analytical purposes have so far concentrated on gaseous samples, which limits potential applications in the field of μTAS and lab-on-a-chip^{47–49}. They are not universal devices able to measure any kind of sample. The main reason for this is the difficulty in achieving adequate sample transport from the liquid to the gas phase, and in increasing the coupled electric power into the plasma without destroying the discharge housing over a certain limit. For miniaturized plasmas, the volume and discharge power is such that even small amounts of liquid can easily extinguish the discharge. For the analysis of liquid samples, microplasmas can either be coupled with different sample-injection devices or applied as plasmas that use one electrode as a liquid, or they can be ignited directly in the liquid⁵⁰. A lot of work is going on in this area and further studies will improve performance to the level of classical analytical plasmas. If microplasmas are to be used with liquids, sample-introduction systems that generate vapours may not always be required. Applications as detectors in a variety of liquid chromatography and capillary electrophoresis modes will require further

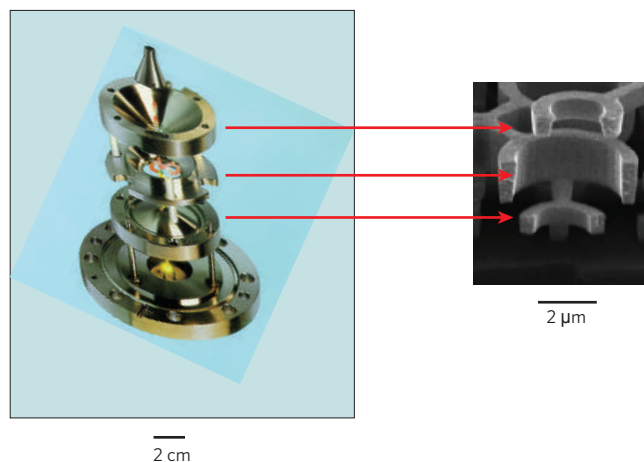


Figure 4 | Ion traps as an example of miniaturization in detection. The left panel shows the ITD700/800, a conventional ion-trap detector (image courtesy of Thermo Electron, Bremen, Germany); the right panel shows a scanning electron micrograph. The arrows indicate the two end caps and the ring electrode. Reprinted, with permission, from ref. 62.

characterization of the role of electrolyte/mobile phase identity on system performance as well. Devices such as these will have a place in the general field of elemental speciation, in which relatively simple devices of low power consumption could be widely applied.

The miniaturization of almost every type of mass analyser (quadrupole ion trap^{51–53}, time of flight^{54–56}, magnetic sector^{57–59} and linear quadrupole^{60,61}) is an active area of research. Miniaturized mass analysers reduce vacuum-system demands, because the maintenance of a constant collision frequency allows an increase in pressure as the analyser size is scaled down. Consequently, power consumption for vacuum and backing pumps may be reduced³⁶. The quadrupole trapping technique (such as the Paul trap) is the most amenable to scaling down to the 1 μm and 10 V regime (Fig. 4). The mass-to-charge ratio of the quadrupole techniques scales proportionally to the applied voltage, inversely with the square of the characteristic length d of the analyser and the angular radio frequency, Ω . A reduction in the trap size requires that either the rf amplitude has to be decreased quadratically (for a constant Ω) or that the Ω has to be increased linearly (for a constant V) with the dimension of the trap, d .

The number of trapped charges in a 1 μm cylindrical ion trap operating at modest voltages will be small⁶². In fact, for trapping times from microseconds to milliseconds during which mass analysis can occur, it is expected that a 1 μm trap will contain at most one ion. Consequently, practical mass spectrometry with 1 μm traps will require a massive array of traps operating in parallel to obtain a useful ion signal.

The future

There has been much interest in the concept of nanofluidics and there seems to be a tremendous potential for more basic research on the subject. However, such devices fall outside the scope of many of the scaling principles described here, and new laws will have to be fully characterized. We believe that nanofluidics has not yet fully found its role. It is often said that there is plenty of room at the bottom, but perhaps not for nanofluidics in the field of traditional analytical chemistry, and not yet for commercialization. Nevertheless, inspired by the nanostructured, biochemically powered machinery of cells, a wealth of possibilities for nanoreactors and transport systems might evolve over the next decade or two.

We have already covered some specific opportunities for future research in the sections above. More broadly, the issue of integration will challenge the chemical engineer: process monitoring and control to rapidly acquire information in combination with additional reaction and preparation steps in one device.

Future research will also need to address some limitations of scaling that have so far prevented the miniaturization of certain devices. An example is high-performance liquid chromatography. Because of pressure drop, today's ultrafast high-performance liquid chromatography techniques demand pumping power of up to 400 bar. Further miniaturization would increase the pressure needed up to thousands of bar. It would not only be hard to construct such powerful pumps, but would also go beyond the mechanical stability of the chip material itself. Therefore, other ways have to be found to move the liquid phase alongside the stationary one. Besides the use of electro-osmotic flow⁶³, Desmet and Baron proposed the use of shear forces and demonstrated the success of this approach for straight channels^{64,65}.

The design of μTAS has required the use of a scientific literature spanning 200 years in a range of fields. This implies that a list of all areas of technology and applications can be consulted to find more missing links. We think that everyone could look at the literature and come up with new ideas for how to develop new approaches to designing miniaturized analysis systems. More approaches are out there, and we will continue to look for them and to exploit them in designing new miniaturized analysis systems to make laboratories more efficient and cost-effective. ■

- Manz, A., Grabner, N. & Widmer, H. M. Miniaturized total chemical analysis systems: a novel concept for chemical sensing. *Sens. Actuators B Chem.* **1**, 244–248 (1990).
- Golay, M. J. E. Vapor phase chromatography and the telegrapher's equation. *Anal. Chem.* **29**, 928–932 (1957).
- van Deemter, J. J., Zuideker, F. J. & Klinkenberg, A. Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. *Chem. Eng. Sci.* **5**, 271–289 (1956).

- Fulford, G. D. & Catchpole, J. P. Dimensionless groups. *Ind. Eng. Chem.* **60**, 71 (1968).
- Knox, J. H. & Gilbert, M. T. Kinetic optimization of straight open-tubular liquid-chromatography. *J. Chromatogr.* **186**, 405–418 (1979).
- Reynolds, O. An experimental investigation of the circumstances which determine whether the motion of water shall be direct or sinuous, and of the law of resistance in parallel channels. *Phil. Trans. R. Soc. Lond.* **174**, 935–982 (1883).
- Drexler, K. E. *Nanosystems: Molecular Machinery, Manufacturing and Computation* (John Wiley & Sons, New York, 1992).
- Trimmer, W. S. N. Microrobots and micromechanical systems. *Sens. Actuators* **19**, 267–287 (1989).
- Manz, A. et al. Planar chips technology for miniaturization and integration of separation techniques into monitoring systems. *J. Chromatogr.* **593**, 253–258 (1992).
- Harrison, D. J. et al. Micromachining a miniaturized capillary electrophoresis-based chemical analysis system on a chip. *Science* **261**, 895–897 (1993).
- Backhouse, C. J., Gajdal, A., Pilarski, L. M. & Crabtree, H. J. Improved resolution with microchip-based enhanced field inversion electrophoresis. *Electrophoresis* **24**, 1777–1786 (2003).
- Ludwig, M., Kohler, F. & Belder, D. High-speed chiral separations on a microchip with UV detection. *Electrophoresis* **24**, 3233–3238 (2003).
- Xu, Y., Zhang, C.-X., Janasek, D. & Manz, A. Sub-second isoelectric focussing in free flow using a microfluidic device. *Lab Chip* **3**, 224–227 (2003).
- Mazereeuw, M., de Best, C. M., Tjaden, U. R., Irth, H. & van der Greef, J. Free flow electrophoresis device for continuous on-line separation in analytical systems. An application in biochemical detection. *Anal. Chem.* **72**, 3881–3886 (2000).
- Kobayashi, H. et al. Free-flow electrophoresis in a microfabricated chamber with a micromodule fraction separator — continuous separation of proteins. *J. Chromatogr. A* **990**, 169–178 (2003).
- Zhang, C.-X. & Manz, A. High-speed free-flow electrophoresis on chip. *Anal. Chem.* **75**, 5759–5766 (2003).
- Fonslow, B. R. & Bowser, M. T. Free-flow electrophoresis on an anodic bonded glass microchip. *Anal. Chem.* **77**, 5706–5710 (2005).
- Kohlheyer, D., Besselink, G. A. J., Schlaumann, S. & Schasfoort, R. B. M. Free-flow zone electrophoresis and isoelectric focusing using a microfabricated glass device with ion permeable membranes. *Lab Chip* **6**, 374–380 (2006).
- Albrecht, J., Gaudet, S. & Jensen, K. F. in *Ninth International Conference on Miniaturized Systems for Chemistry and Life Sciences* (eds Jensen, C. F., Han, J., Harrison, D. J. & Voldman, J.) 1537–1539 (Transducer Research Foundation, Boston, Massachusetts, 2005).
- Janasek, D., Schilling, M., Franzke, J. & Manz, A. Isotachophoresis in free-flow using a miniaturized device. *Anal. Chem.* **78**, 3815–3819 (2006).
- Tas, N. R., Mela, P., Kramer, T., Berenschot, J. W. & van den Berg, A. Capillary induced negative pressure of water plugs in nanochannels. *Nano Lett.* **3**, 1537–1540 (2003).
- Eijkel, J. C. T. & van den Berg, A. Nanofluidics: what is it and what can we expect from it? *Microfluid. Nanofluid.* **1**, 249–267 (2005).
- Eijkel, J. C. T. & van den Berg, A. The promise of nanotechnology for separation devices — from a top-down approach to nature-inspired separation devices. *Electrophoresis* **27**, 677–685 (2006).
- Bessoth, F. G., deMello, A. J. & Manz, A. Microstructure for efficient continuous flow mixing. *Anal. Commun.* **36**, 213–215 (1999).
- Pollack, L. et al. Compactness of the denatured state of a fast-folding protein measured by submillisecond small-angle X-ray scattering. *Proc. Natl Acad. Sci. USA* **96**, 10115–10117 (1999).
- Stroock, A. D. et al. Chaotic mixer for microchannels. *Science* **295**, 647–651 (2002).
- Song, H., Bringer, M. R., Tice, J. D., Gerds, C. J. & Ismagilov, R. F. Experimental test of scaling of mixing by chaotic advection in droplets moving through microfluidic channels. *Appl. Phys. Lett.* **83**, 4664–4666 (2003).
- Lee, C.-C. et al. Multistep synthesis of a radiolabeled imaging probe using integrated microfluidics. *Science* **310**, 1793–1796 (2005).
- Lipshutz, R. J., Fodor, S. P. A., Gingeras, T. R. & Lockhart, D. J. High density synthetic oligonucleotide arrays. *Nature Genet.* **21**, 20–24 (1999).
- Fodor, S. P. A. et al. Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**, 767–773 (1991).
- Dittrich, P. S. & Manz, A. Single-molecule fluorescence detection in microfluidic channels — the Holy Grail in μTAS ? *Anal. Bioanal. Chem.* **382**, 1771–1782 (2005).
- Lakowitz, J. R. (ed.) *Principles of Fluorescence Spectroscopy* (Kluwer, New York, 1999).
- Kitamori, T., Tokeshi, M., Hibara, A. & Sato, K. Thermal lens microscopy and microchip chemistry. *Anal. Chem.* **76**, 52A–60A (2004).
- de Mello, A. J. Seeing single molecules. *Lab Chip* **3**, 29N–34N (2003).
- Keller, R. A. et al. Analytical applications of single-molecule detection. *Anal. Chem.* **7**, 316A–324A (2002).
- Palmer, P. T. & Limero, T. F. Mass spectrometry in the US space program: past, present, and future. *J. Am. Soc. Mass Spectrom.* **12**, 656–675 (2001).
- Short, R. T., Fries, D. P., Koler, S. K., Lembke, C. E. & Byrne, R. H. Development of an underwater mass-spectrometry system for *in situ* chemical analysis. *Meas. Sci. Technol.* **10**, 1195–1201 (1999).
- Patterson, G. E. et al. Miniature cylindrical ion trap mass spectrometer. *Anal. Chem.* **74**, 6145–6153 (2002).
- Ritter, L. S. et al. Analytical performance of a miniature cylindrical ion trap mass spectrometer. *Anal. Chem.* **74**, 6154–6162 (2002).
- Townsend, J. S. *Electricity in Gases* 365 (Clarendon, Oxford, 1915).
- Paschen, F. Über die zum Funkenübergang in Luft, Wasserstoff und Kohlensäure bei verschiedenen Drücken erforderliche Potentialdifferenz. *Wied. Anal. Phys. Chem.* **37**, 69–96 (1889).
- Holm, R. Der gegenwärtige Stand der Theorie des Glimmstroms. *Phys. Z.* **25**, 497–535 (1924).
- von Engel, A. & Steenbeck, M. *Elektrische Gasentladungen* Vol. II (Springer, Berlin, 1934).
- Margenau, H. Theory of high frequency gas discharges. IV. Note on the similarity principle. *Phys. Rev.* **73**, 326–328 (1948).

45. Llewellyn Jones, F. & Morgan, G. D. High-frequency discharges. I. Breakdown mechanism and similarity relationship. *Proc. Phys. Soc. Lond. B* **64**, 560–573 (1951).
46. Encyclopedia of Physics Vol. XXII *Gas Discharges II* (ed. Flüge, S.) 1–40 (Springer, Berlin, 1956).
47. Franzke, J., Kunze, K., Miclea, M. & Niemax, K. Microplasmas for analytical spectrometry. *J. Anal. Atom. Spectrom.* **18**, 802–807 (2003).
48. Miclea, M., Kunze, K., Franzke, J. & Niemax, K. Microplasma jet mass spectrometry of halogenated organic compounds. *J. Anal. Atom. Spectrom.* **19**, 990–994 (2004).
49. Franzke, J. & Miclea, M. Sample analysis with miniaturized plasmas. *Appl. Spectrosc.* **60**, 80A–90A (2006).
50. Jenkins, G., Franzke, J. & Manz, A. Direct optical emission spectroscopy of liquid analytes using an electrolyte as a cathode discharge source (ELCAD) integrated on a micro-fluidic chip. *Lab Chip* **5**, 711–718 (2005).
51. Riter, L. S., Laughlin, B. C., Nikolaev, E. & Cooks, R. G. Direct analysis of volatile organic compounds in human breath using a miniaturized cylindrical ion trap mass spectrometer with a membrane inlet. *Rapid Commun. Mass Spectrom.* **16**, 2370–2373 (2002).
52. Badman, E. R. & Cooks, R. G. Cylindrical ion trap array with mass selection by variation in trap dimensions. *Anal. Chem.* **72**, 5079–5086 (2000).
53. Moxom, J., Reilly, P. T. A., Whitten, W. B. & Ramsey, J. M. Analysis of volatile organic compounds in air with a micro ion trap mass analyzer. *Anal. Chem.* **75**, 3739–3743 (2003).
54. Berkout, V. D., Cotter, R. J. & Segers, D. P. Miniaturized EI/Q/oa TOF mass spectrometer. *J. Am. Soc. Mass Spectrom.* **12**, 641–647 (2001).
55. Prieto, M. C., Kovtoun, V. V. & Cotter, R. J. Miniaturized linear time-of-flight mass spectrometer with pulsed extraction. *J. Mass Spectrom.* **37**, 1158–1162 (2002).
56. Cornish, T. J., Ecelberger, S. & Brinckerhoff, W. Miniature time-of-flight mass spectrometer using a flexible circuitboard reflector. *Rapid Commun. Mass Spectrom.* **14**, 2408–2411 (2000).
57. Diaz, J. A., Giese, C. F. & Gentry, W. R. Sub-miniature ExB sector-field mass spectrometer. *J. Am. Soc. Mass Spectrom.* **12**, 619–632 (2001).
58. Diaz, J. A., Giese, C. F. & Gentry, W. R. Portable double-focusing mass-spectrometer system for field gas monitoring. *Field Anal. Chem. Technol.* **5**, 156–167 (2001).
59. Diaz, J. A., Giese, C. F. & Gentry, W. R. Mass spectrometry for *in-situ* volcanic gas monitoring. *Trends Anal. Chem.* **21**, 498–514 (2002).
60. Boumsellek, S. & Ferran, R. J. Trade-offs in miniature quadrupole designs. *J. Am. Soc. Mass Spectrom.* **12**, 633–640 (2001).
61. Orient, O. J. & Chutjian, A. A compact, high-resolution Paul ion trap mass spectrometer with electron-impact ionization. *Rev. Sci. Instrum.* **73**, 2157–2160 (2002).
62. Blain, M. G. *et al.* Towards the hand-held mass spectrometer: design considerations, simulations, and fabrication of micrometer-scaled cylindrical ion trap. *Int. J. Mass Spectrom.* **236**, 91–104 (2004).
63. Bruin, G. J. M., Tock, P. P. H., Kraak, J. C. & Poppe, H. Electrically driven open-tubular liquid chromatography. *J. Chromatogr.* **517**, 557–573 (1990).
64. Desmet, G. & Baron, G. V. On the possibility of shear-driven chromatography: a theoretical performance analysis. *J. Chromatogr. A* **855**, 57–70 (1999).
65. Desmet, G. & Baron, G. V. The possibility of generating high-speed shear-driven flows and their potential application in liquid chromatography. *Anal. Chem.* **72**, 2160–2165 (2000).
66. de la Rue, W. & Muller, H. W., Experimental researches on the electric discharge with the chloride of silver battery. *Phil. Trans. R. Soc. Lond.* **171**, 65–116 (1880).
67. Carr, W. R. On the laws governing electric discharges in gases at low pressures. *Phil. Trans. R. Soc. Lond. A* **201**, 403–433 (1903).

Acknowledgements We thank D. Agar for proof reading. The financial support of the Ministerium für Innovation, Wissenschaft, Forschung und Technologie des Landes Nordrhein-Westfalen, of the Bundesministerium für Bildung und Forschung, and of the Deutsche Forschungsgemeinschaft (D.F.G.) is gratefully acknowledged.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to D.J. (d.janasek@ansci.de).

Developing optofluidic technology through the fusion of microfluidics and optics

Demetri Psaltis¹, Stephen R. Quake² & Changhuei Yang¹

We describe devices in which optics and fluidics are used synergistically to synthesize novel functionalities. Fluidic replacement or modification leads to reconfigurable optical systems, whereas the implementation of optics through the microfluidic toolkit gives highly compact and integrated devices. We categorize optofluidics according to three broad categories of interactions: fluid–solid interfaces, purely fluidic interfaces and colloidal suspensions. We describe examples of optofluidic devices in each category.

Optofluidics refers to a class of optical systems that are synthesized with fluids. Fluids have unique properties that cannot be found in solid equivalents, and these properties can be used to design novel devices. Examples of such properties include: the ability to change the optical property of the fluid medium within a device by simply replacing one fluid with another; the optically smooth interface between two immiscible fluids; and the ability of flowing streams of miscible fluids to create gradients in optical properties by diffusion. Most optical systems are currently made with solid materials such as glasses, metals and semiconductors, but there are cases in which it has been advantageous to use fluids. The oil-immersion microscope¹, liquid mirrors for telescopes², liquid-crystal displays³ and electrowetting lenses⁴ are good examples. Here we describe the various methods used to implement optofluidic devices with recently developed microfluidic technologies that allow accurate control of liquids on small spatial scales. Integration and reconfigurability are two major advantages associated with optofluidics. Whereas microfluidics has made it possible to integrate multiple fluidic tasks on a chip, most optical components, such as the light source, sensors, lenses and waveguides, remained off the chip. Optofluidic integration combines optics and microfluidics on the same chip by building the optics out of the same fluidic toolkit. The second advantage of optofluidics lies in the ease with which one can change the optical properties of the devices by manipulating fluids.

Microfluidics is a burgeoning field with important applications in areas such as biotechnology, chemical synthesis and analytical chemistry. Many of these applications of microfluidics are discussed elsewhere in this issue, and for the purposes of this Review we emphasize only a few salient points. First, there is now an extensive body of literature on how the physical properties of fluids that are available only on small spatial scales can be exploited for device functionality^{5–7}. Many of these effects can also be used to control optical properties. Second, technological advances in device fabrication have made it possible to build miniaturized devices with complex networks of channels, valves, pumps and other methods of fluidic encapsulation and manipulation⁸. This creates a powerful set of tools for fluidic control, and because the feature sizes in these systems are shrinking over time, they will inevitably approach the wavelength of visible light. When this convergence happens, optofluidic-device applications will increase dramatically, just as the ability to lithographically define features with dimensions just below the wavelength of light has created a revolution in photonics. Third, a large anticipated market for microfluidics lies in portable devices for environmental monitoring, medical diagnostics

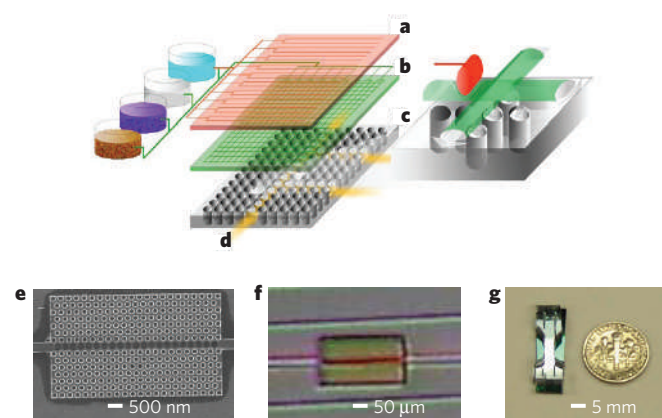


Figure 1 | A generalized layer construction of an optofluidic device. An optofluidic device typically consists of three layers. **a**, The topmost layer consists of the microfluidic controls. Microfluidic valves and pumps may be incorporated in this layer. **b**, The middle layer contains the microfluidic channels. **c**, The third layer is the optical structure and may contain photonic crystal structures, sensors, sources and waveguides. **d**, Light can be guided within the third layer. **e–g**, The range of scales²³ involved in the fabricated device will typically span from subwavelength (~tens to hundreds of nanometres) for the optical structures to tens and hundreds of millimetres for the microfluidic structures.

and chemical-weapon detection. Optofluidics is helping to realize these aspirations by combining optical elements into microfluidic devices in ways that increase portability and sensitivity.

Although any microfluidic fabrication method can, in principle, be adapted to synthesize optofluidic devices, most of the implementations thus far have been with soft lithography^{9–11} (Fig. 1). Soft lithography allows rapid fabrication of complex microfluidic structures¹² in flexible polymer substrates at a fraction of the cost of traditional glass or semiconductor manufacturing. Although the most popular polymer is poly(dimethylsiloxane) (PDMS), many other materials are suitable, such as photocurable hydrogels¹³, thermoset plastics¹⁴ and elastomers¹⁵, and photocurable solvent-resistant elastomers such as perfluoropolyethers (PFPE)¹⁶. The optical transparency and good optical quality of PDMS has been demonstrated in applications such as soft lithographic fabrication of blazed gratings¹⁷ and solid immersion lenses¹⁸, showing that these materials are suitable in the optofluidic context.

¹Department of Electrical Engineering, California Institute of Technology, Pasadena, California 91125, USA. ²Department of Bioengineering, Stanford University and Howard Hughes Medical Institute, Stanford, California 94305, USA.

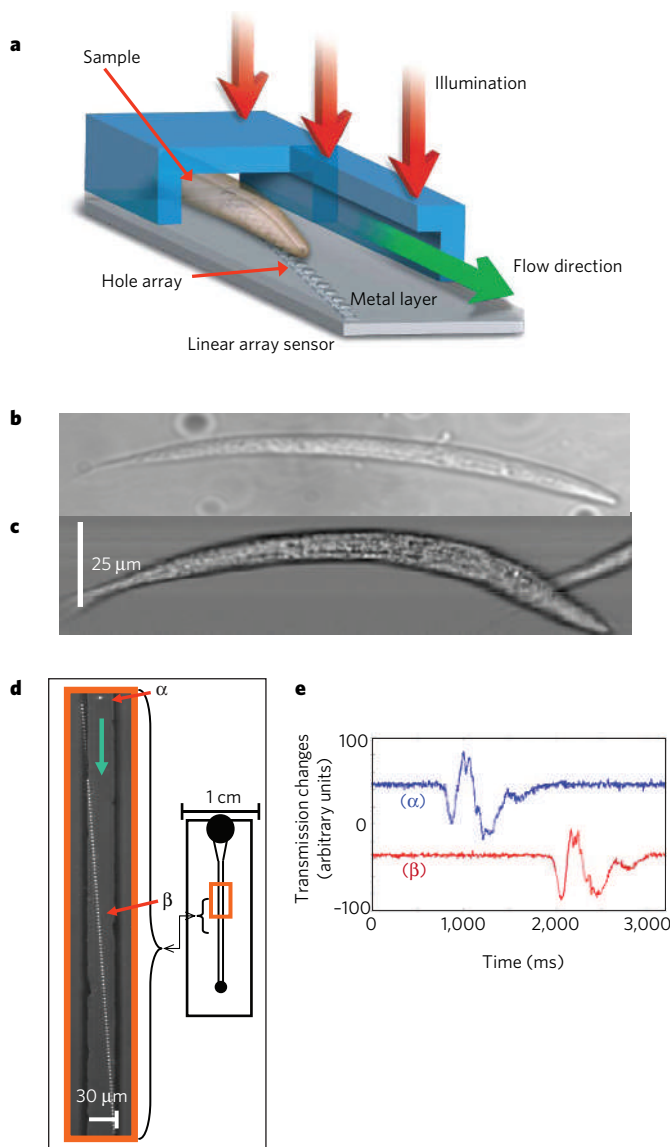


Figure 2 | The optofluidic microscope. **a**, Implementation scheme for an on-chip optofluidic microscope. The device is uniformly illuminated from the top. The target sample flows through the channel, and the transmission through each hole is acquired and recorded. The composition of the transmission traces creates a transmission image of the target sample. **b**, A conventional microscope image of *Caenorhabditis elegans*. **c**, An optofluidic microscope (700-nm resolution) image of *C. elegans*. **d**, By staggering the holes along the length of the channel, the separation between holes can be made equal to the pixel size of the underlying sensor array and enable the unique mapping of each hole to a pixel. The lateral displacement of the holes across the channel can be made arbitrarily small and it defines the resolution of the microscope. This approach enables the construction of microscopes with resolutions that are much finer than the pixel resolution of a conventional sensor grid. **e**, The transmission trace through two representative holes, α and β , on the microscope as the sample flows across them.

Optofluidic devices can be categorized into three major groups: structured solid–liquid hybrids in which the optical properties of both media are relevant; complete fluid-based systems in which only the optical properties of the fluids are relevant; and colloid-based systems in which manipulation of solid particles in liquid, or using the unique optical properties of colloidal solution, form the basis of the optofluidic devices. Recent progress in this last discipline has led to novel nanoparticles with interesting optical properties along with methods to control them in the solution phase.

Fluids in solids

The interface between materials with differing optical properties is commonly used for controlling the propagation of light. For example, the interface between air and a dispersive material such as glass in a prism separates the colours of an incident beam into different angles. Another important example can be found in fibre-optic waveguiding, in which total internal reflection at the interface between the cladding and a core with a higher index of refraction confines light to the core.

Interfaces between solid and liquid materials can also be used to synthesize optofluidic devices. This can be done by fabricating a solid structure with voids whose dimensions are larger than the wavelength of light. The reflection and transmission of light through the void is then modified by the insertion of liquids in the voids. Microfluidics and microfabrication provide the tools for making such devices, of which the Agilent switch¹⁹ is one of the earliest examples. In this case, inkjet technology was used to introduce or remove an index matching liquid from the back of a total internal reflection (TIR) mirror. The transmission of light could thus be switched on and off. More recently, Campbell *et al.*²⁰ demonstrated an exchange-bypass switch using the same principle. Adaptive lenses can be formed in this way by varying the fluidic pressure in a soft PDMS encapsulation^{21,22}. Finally, a periodic array of voids, in which each void can be selectively filled with absorbing dye, can be used to write an arbitrary two-dimensional pattern. This can therefore serve as a display or memory device. Such a device has been demonstrated by Quake's group using an extensive microfluidic circuit of channels and valves¹².

Any adaptive device can be thought of as a sensor if the adaptation is due to something one wishes to characterize. In this regard, optofluidic adaptation mechanisms can often be used as sensors. For example, the compact tunable microfluidic interferometer described in ref. 23, in which a fluid–air interface is used to provide an optical path difference in a compact interferometer, can be adapted for biosensing or chemical analysis. The optofluidic microscope (OFM) is another good example²⁴ (Fig. 2). Compiling the time-dependent changes of the optical transmission across a fluidic channel gives one the image of the sample flowing in the channel. The fact that the function of a microscope can be integrated into an optofluidic chip implies that multiple sensors can also be packed onto a compact chip and parallel processing of the sample be performed. Other examples of optofluidic sensors can be found in the compact resonant integrated microfluidic refractometer²⁵ and fibre Bragg grating refractive-index sensors²⁶.

A phenomenon related to total internal reflection, known as evanescent waves, also has applications in optofluidic devices. Evanescent waves are launched in the region of lower index when light undergoes total internal reflection at an interface. Even though evanescent modes cannot carry energy away from the interface, light propagation in the region of higher index can be strongly affected by the evanescent modes. This provides a non-intrusive mechanism for optofluidic control by introducing liquids with specific optical properties into the region surrounding the optical structure. For example, the phase velocity of a dielectric waveguide depends on the refractive index of the cladding. By selecting the index of the liquid that is inserted in the region surrounding the dielectric waveguide, we can adjust the phase delay through the waveguide. This has many applications in integrated optics, including optical switching and modulation. For instance, the adjustable phase delay has recently been used to tune a ring resonator that is constructed as a closed loop of a dielectric waveguide²⁷ (Fig. 3). Besides phase-delay changes, the introduction of an absorbing liquid (a dye, for example) introduces attenuation of the light propagating in the waveguide. Optofluidic devices that rely on evanescent mode modification can also function as sensors. Surface-plasmon-based sensors^{28–30} are a well-known example of this. Other recent examples in this category include the optofluidic sensors based on the high-Q resonators developed by Vahala's group³¹ and the zero-mode waveguide sensors³².

In our discussion so far, we have assumed that the size of the voids is significantly larger than the optical wavelength. When the size of the voids is approximately equal to the wavelength, the effect of liquid insertion

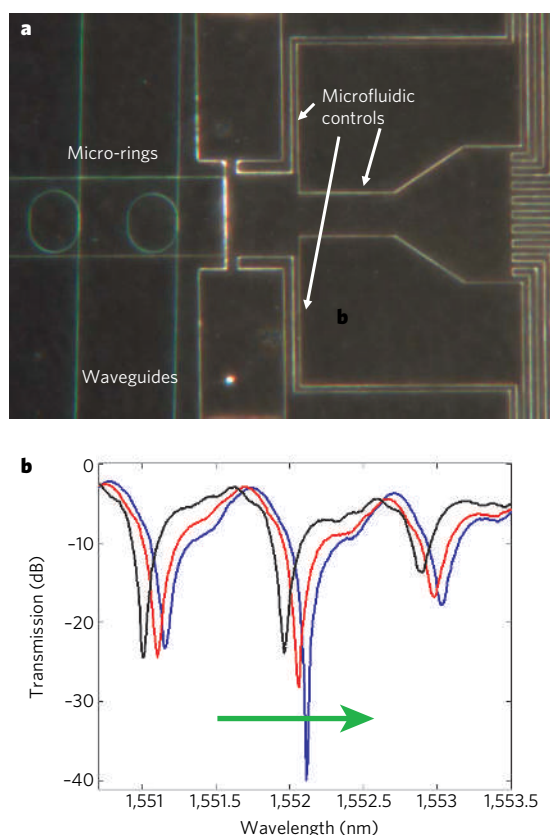


Figure 3 | Optical filter based on an optofluidic micro-ring resonator. **a**, The optical structure consists of a bus waveguide coupled to a micro-ring waveguide resonator. Liquid in the microchannel constitutes the upper cladding of the waveguides. The refractive index of the liquid controls the resonance wavelength and the strength of the coupling. **b**, At resonance, light travelling in the waveguide will be strongly attenuated. The graph shows that this resonance can be tuned by changing the refractive index of the liquid. The arrow indicates the increase in resonance wavelength as the refractive index increases. (Reprinted, with permission, from ref. 27.)

can become more complex. For instance, a blazed grating immersed in liquid-crystal medium can be turned on and off by modifying the index of the liquid³³. As another example, consider a single-mode optical waveguide constructed by inserting liquid into a rectangular channel of the appropriate size. It would guide light in the liquid when the index of the liquid exceeds that of the surrounding medium. Such a liquid-core optofluidic waveguide is useful for measuring the properties of small volumes of liquid samples, because a long or highly structured waveguide can be easily realized³⁴. The idea can be further extended by introducing a periodic structure on the walls of the waveguide with period Λ . This periodic structure strongly reflects light with wavelengths equal to $2n\Lambda/m$, where m is a positive integer and n is the index of refraction for the liquid. The frequency response of this filter can thus be tuned by modifying the index of the liquid. Such a structure has been used to show a distributed feedback dye laser³⁵ (Fig. 4) — one of several optofluidic dye lasers that have been demonstrated recently^{36–38}. As a class of optofluidic devices, these lasers are useful on their own and constitute an important set of enabling optofluidic platform tools. Among other advantages, these lasers are highly compact, widely tunable and robust. Their pump sources can be either an external light illumination or an on-chip laser diode illumination.

Photonic band gap (PBG) structures³⁹ are particularly interesting for optofluidics because they naturally have voids into which fluids can be injected. A completely uniform PBG structure is of little practical interest. It is the introduction of 'defects' that gives PBGs interesting functionalities, such as waveguiding capabilities or resonances. A defect consists of one or several voids in the periodic structure; typically, the size of the void varies and is defined lithographically when the device is fabricated. In the context of optofluidics, defects can be easily created or reconfigured by filling specific voids with liquid⁴⁰. PBG structures can thus be reconfigured optofluidically by interfacing the photonic and microfluidic circuits. An integrated fluidic/photonic device that fluidically switches a PBG waveguide⁴¹ has recently been demonstrated (Fig. 1). In addition, a PBG cavity can be formed by a single defect. The resonant frequency of the PBG structure can be tuned by the liquid inserted at the defect site⁴². Such cavities have been used to build PBG lasers^{43,44}, sensors⁴⁵ and slow-light structures⁴⁶. PBG fibres⁴⁷ are of great interest for optofluidics as well, because the cladding or the core can contain hollow regions into which fluids can be inserted, thereby modifying a fibre's transmission properties^{34,48}.

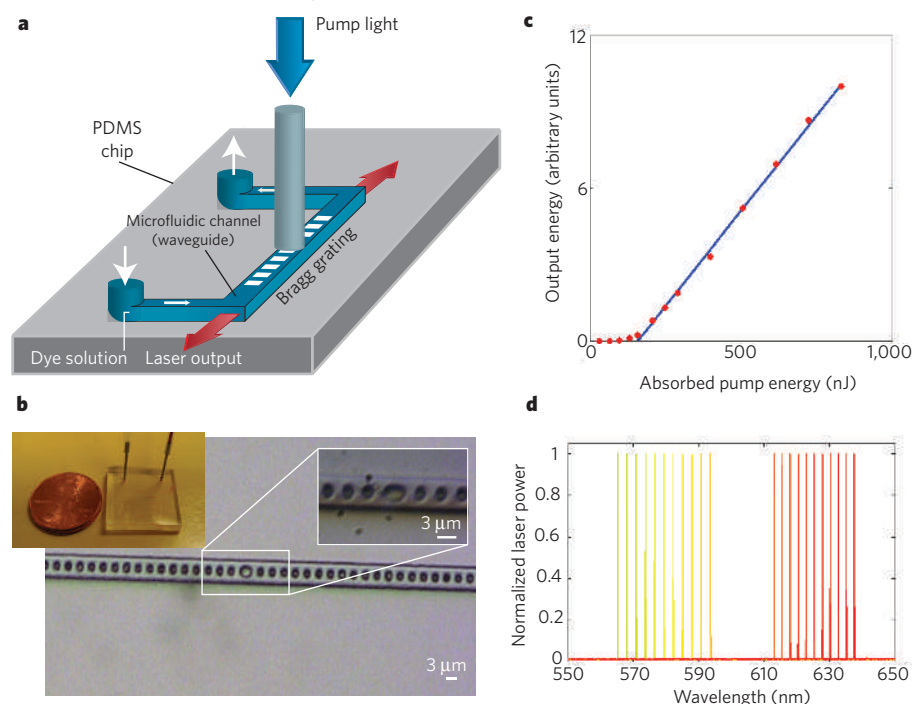


Figure 4 | An optofluidic distributed feedback (DFB) dye laser. **a**, A schematic of the laser. A microfluidic channel with a distributed feedback Bragg grating structure is fabricated in PDMS. A dye solution, which acts as both the core of the optical waveguide and the gain medium, can be introduced into the structure through the channel. An excitation pump light field is incident on the laser structure. **b**, An optical micrograph of the DFB structure. The laser's wavelength may be tuned by changing the dye choice or by stretching/compressing the flexible PDMS to change the periodicity of the DFB structure. Insert shows a device next to a penny to indicate its small size. **c**, Experimental data demonstrating the threshold characteristic of the laser. **d**, Demonstration of ~60 nm tuning range for the laser that can be achieved by mechanically deforming the structure and changing the dye choice (yellow, Rhodamine 6G dye; red, Rhodamine 101 dye).

Looking forward into the near future we can expect to see optofluidic systems that are based on porous solid systems in which sub-wavelength voids or pores are randomly distributed. In such systems, the voids do not scatter light but collectively modify the effective index of the homogeneous material. Insertion of a liquid into the voids can tune the refractive index, absorption and birefringence⁴⁹ of such systems.

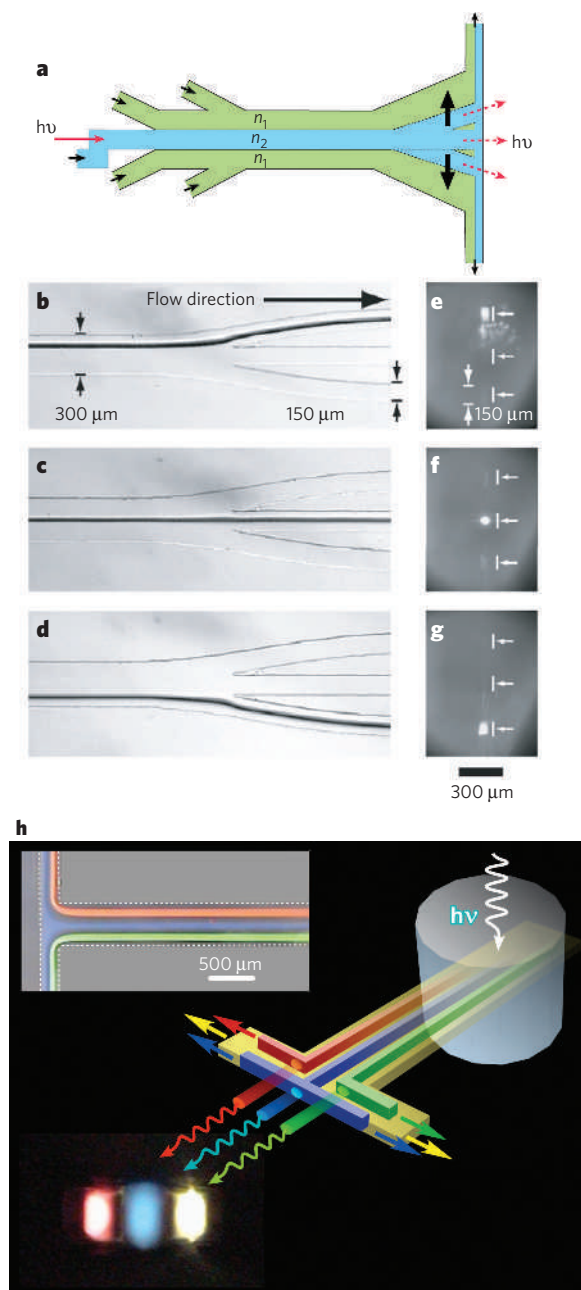


Figure 5 | The liquid-core/liquid-cladding (L^2) waveguide. **a**, Schematic of an L^2 waveguide. The green and blue regions represent two different liquids. By choosing the inner-core liquid to be of higher refractive index (n_2) than the refractive index (n_1) of the surrounding media, light can be guided within the inner core. The red arrows indicate the direction of light input and propagation. $h\nu$ is the energy quantum of the light. **b–d**, The direction of light propagation can be easily altered by altering the differential flow rates for the liquids to steer the inner-core liquid. **e–g**, Experimental verification of light output steering via fluid-core steering. **h**, By channelling different fluorescent dyes in the waveguide, it is possible to create microfluidically controllable light sources. (Reprinted, with permission, from ref. 67.)

Fluids in fluids

Our discussion thus far has focused on the use of different combinations of solids and liquids to elicit changeable optical effects. The replacement of the solid substrate with a liquid in an optofluidic system can result in an even greater degree of flexibility. Whitesides' group's work on liquid-core/liquid-cladding (L^2) optical waveguides is an excellent example of such a system⁵⁰. An L^2 waveguide typically consists of a microfluidic channel in which an optically dense fluid flows within an enveloping sheath of fluid with a lower refractive index (Fig. 5). In addition to serving as waveguides, this method can also be used to create tunable light sources by the addition of dyes to the core fluid⁵¹.

Purely fluidic systems have several unique advantages that cannot be obtained from solid–fluid hybrids. First, the physical profile of the fluidic system that is optically relevant can be quite independent of the actual profile of the underlying microfluidic system. For example, in the case of the L^2 waveguide, light travelling through this dynamic waveguide can be directed and steered by simply changing the relative flow profile and flow rate of the fluid⁵⁰. Second, and perhaps more importantly, it is possible to achieve finer features in the composite fluid media. For example, ref. 50 reported the establishment of a $<10\ \mu\text{m}$ jet stream that was formed in a large ($>100\ \mu\text{m}$) microfluidic channel. Finally, the interface between the two fluids can be made optically smooth by either choosing two immiscible fluids or by flowing the fluids next to each other at low Reynolds numbers. The roughness of the underlying microfluidic architecture has little impact on the fluid interfaces.

It has long been recognized that the optical smoothness of fluid interfaces can be a useful and cost-effective way to circumvent the challenges of creating surfaces of similar quality in solid systems. Perhaps even more interesting in the optical context is the fact that the meniscus between two immiscible fluids of equal density in a column is perfectly spherical — a curvature profile that is used in most commercially available lenses. An example of how this optically relevant interface has been made use of can be found in the variable-focus liquid lens that was developed by Philips Research Eindhoven⁴. The lens consists of two immiscible liquids of different refractive indices placed in a cylindrical housing. The lens is designed in such a way that the curvature of the meniscus and thus the effective focal length of the lens can be altered by electrowetting⁵². The same physical mechanism has also been used for displays by modifying the transmittance of a pixel through the movement of a dye solution⁵³.

Miscible liquids and their interfaces can also be of significant use in the optofluidic context⁵⁴. The diffusion across the interface of two liquids is another unique property that doesn't have a direct equivalent in solid-based devices. Specifically, the diffusion process can create a concentration and refractive-index gradient. The controllability and flexibility by which this gradient can be adjusted through flow parameters and fluid choices enable the creation of novel optical interconnects. For example, an optical splitter and wavelength filter based on the selective mixing of two fluid jets in a third fluidic medium has been demonstrated⁵⁵. Unlike a conventional beamsplitter, the split ratio of the optofluidic beamsplitter can be dynamically tuned for any given wavelength by changing the absorbance properties of the input fluidic jets.

Solids in fluids

There is a third possibility that is optically interesting: the introduction of solid particles in a fluid. Such colloidal optofluidic systems can enable large changes in the optical properties of the resulting medium. Furthermore, small particles or colloids within the fluid can be manipulated with relative ease and precision, thereby creating gradients in the optical properties or very localized modifications. The ramifications of such manipulations in terms of synthesizing novel optical functionality are wide ranging and exploration of these has only begun recently.

Small particles are optically interesting in various ways. For example, dielectric particles of size comparable to the wavelength can exhibit significant spectral and angular scattering variations⁵⁶. Particles that are significantly smaller than the wavelength can be used to change the refractive index or absorption coefficient of the medium as a whole⁵⁶.

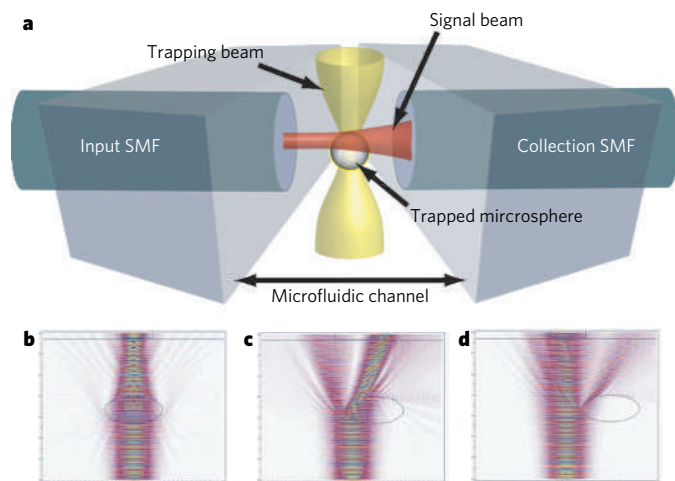


Figure 6 | The all-optical switch based on optofluidic beam manipulation. **a**, Schematic of the switch. The trapping beam steers a microsphere between two single-mode fibres (SMFs). The extent of coupling between the two fibres depends on the position of the microsphere. On the basis of its position, the microsphere can either focus the light from one fibre into the next, or deflect it away. **b–d**, Simulation illustrating that the beam deflection depends on the microsphere displacement.

Particles that are significantly larger than the wavelength can be treated as a spherical focusing lens in the medium^{57,58}. Quantum dots can be designed to fluoresce or scatter with dramatically improved efficiency⁵⁹. It is also interesting to note that quantum dots can be designed to fluoresce over a very narrow ($\sim$ tens of nanometres) and pre-determined spectral bandwidth. Recently developed crescent-shaped nanoparticles have enhanced Raman scattering cross-sections⁶⁰ and other optically interesting properties⁶¹.

There are several methods for optically manipulating the particles. Optical tweezing, one of the more established and popular methods, has seen significant improvements over the past decade. At present, up to 400 particles can be simultaneously manipulated^{62,63}. In addition to optical tweezing, particles can be transported in a fluid through electric fields if they are charged, or through electric field gradients if they have a different electric permittivity from the surrounding medium. Fluidic flows

can also be used to transport and affect the particle concentration. Finally, if ferromagnetic layers are included in the construction of the particles, magnetic fields can control their orientation and trajectory. A particularly interesting method was recently developed in which both light and electric fields are combined to effect particle manipulation⁶⁴. In this approach, an optical image is projected onto a specially prepared glass substrate. The light field activates the coated photoconductive layer and creates non-uniform electric fields. These fields can either attract or repel particles that are in the solution above the substrate. This approach of optofluidic particle manipulation is especially noteworthy in terms of the number of particles that can be manipulated simultaneously — 15,000 traps were created over an area of about 1 mm² in the demonstration experiment.

The ability to optically control particles can be used to create a range of novel optofluidic devices. A clear and elegant example is the optofluidic beam manipulator demonstrated by Dumachuk *et al.*⁵⁷. The general concept is as follows. An optically trapped microsphere that is placed in front of the exit port of a waveguide can function as an optically movable lens for beam manipulation. By steering the microsphere, the output beam can be deflected in a range of directions. Dumachuk *et al.*⁵⁶ used the method to create an all-optical switch by steering the microsphere between two mutually facing waveguides that are separated by a short distance (Fig. 6). The transmission of light from one waveguide to the other is enhanced when the microsphere is well centred and functions as a focussing lens. Individual particles can also be used to store bits of information. In a recent demonstration, a nanoparticle memory was synthesized by electrically transporting quantum dots to selected sites in a substrate. The spectral composition of the quantum-dot cocktails represents the stored information (Fig. 7).

The simple presence of the particles in a solution modifies its optical properties. This effect was used to demonstrate a liquid–liquid waveguide in which the core is a colloidal suspension surrounded by a cladding of pure solvent. By changing the diameter and volume fraction of the particles, it was possible to alter the propagation properties of the waveguide. The presence of nanoparticles in a liquid also affects the nonlinear optical properties of the medium. For example, the plasmon resonance of metal nanoparticles enhances the field locally by two or three orders of magnitude. The increase in the field strength leads to an enhancement of the nonlinear response of the medium, a property that has been used for surface-enhanced Raman scattering. Thermal nonlinearities introduced by metal nanoparticles in a solvent were recently used to demonstrate holographic recording in a fluid⁶⁵.

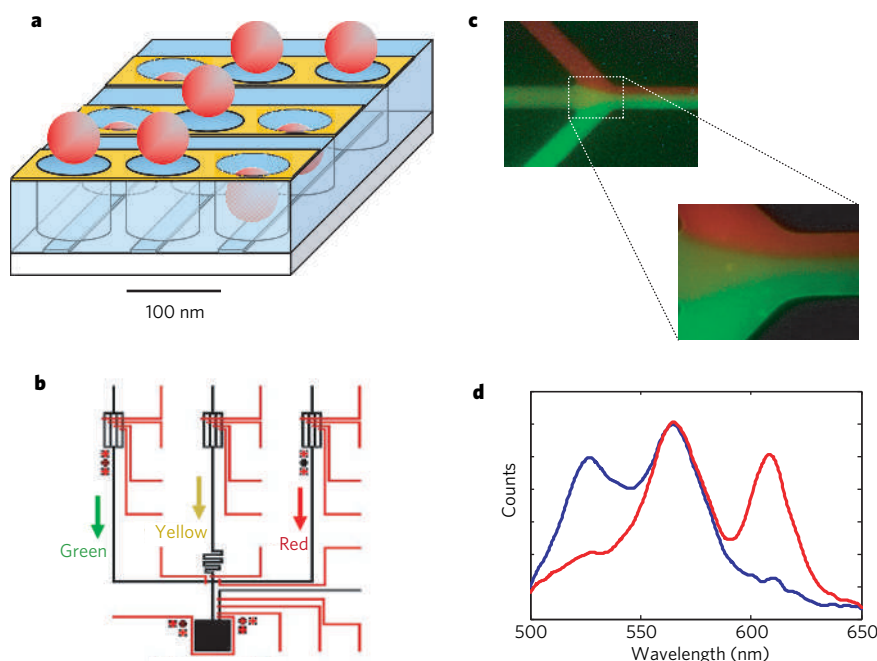


Figure 7 | An optofluidic memory system based on nanowells and quantum dots. **a**, An array of electrically addressable nanowells is used to trap quantum dots with resolution smaller than an optical wavelength. Data can be encoded by the presence or absence of photoluminescence in each region, or by spectral codes consisting of cocktails of quantum dots with distinct emission spectra in a single region. **b**, The microfluidic mixing circuit used to supply multiple quantum dots with different photoluminescence wavelengths to the nanowell array. **c**, A fluorescence image of the mixing of three species of quantum dots. Here the flow rate of each channel can be controlled to vary the relative concentrations of the types of dot, to generate distinct spectral codes. **d**, A spectrogram of two separate codes produced by the mixing circuit. Each peak in the spectrum represents a single base M digit, where M is the number of distinguishable intensity levels. A memory with N species of quantum dots gives an information density of $N \log_2(M)$ bits per nanowell. (Reprinted, with permission, from ref. 68.)

The ability to optically manipulate colloidal particles can be used as a mechanism for optically controlling the fluidic flow. An example of this is the recent work by Lee's group⁶⁶. In this work, the enhanced optical absorption and subsequent heat generation of metallic nanoparticles is used. The method is simple and direct. Nanoparticles are added to the liquid of interest. A light beam is then introduced at the air–liquid interface. Due to the enhanced optical absorption and heat generation, the liquid evaporates and condenses beyond the edge. Subsequent condensation of the droplets wets the region adjacent to the air–liquid interface and causes a shift of the entire liquid body.

Looking forward

We conclude by noting that optofluidics is a field still in its infancy. It is full of promise, but few devices have been successfully commercialized as of yet. Many technological elements and even some fundamental concepts are just now being developed. As technological tools, optofluidic devices are good complements to micro-electromechanical systems (MEMS). Where MEMS physically actuate solid components to achieve reconfigurability, optofluidic devices reconfigure themselves through fluidic controls. Finally, as we have described in this Review, the synergy between optics and fluids is particularly rich. One currently understudied area that is especially noteworthy is the optical control of fluids. Such controls can dramatically simplify fluid manipulation by completely removing the need for on-chip plumbing. ■

- Ronchi, V. Giovan Battista Amici's contribution to the advances of optical microscopy. *Physica* **11**, 520–533 (1969).
- Wood, R. W. *Astrophys. J.* The mercury paraboloid as a reflecting telescope. **29**, 164–176 (1909).
- Haas, W. E. Liquid-crystal display research — the first 15 years. *Mol. Cryst. Liq. Cryst.* **94**, 1–31 (1983).
- Kuiper, S. & Hendriks, B. H. W. Variable-focus liquid lens for miniature cameras. *Appl. Phys. Lett.* **85**, 1128–1130 (2004).
- Squires, T. M. & Quake, S. R. Microfluidics: fluid physics at the nanoliter scale. *Rev. Mod. Phys.* **77**, 977–1026 (2005).
- Stone, H. A., Stroock, A. D. & Ajdari, A. Engineering flows in small devices: microfluidics toward a lab-on-a-chip. *Annu. Rev. Fluid Mech.* **36**, 381–411 (2004).
- Atencia, J. & Beebe, D. J. Controlled microfluidic interfaces. *Nature* **437**, 648–655 (2005).
- Hong, J. W. & Quake, S. R. Integrated nanoliter systems. *Nature Biotechnol.* **21**, 1179–1183 (2003).
- Duffy, D. C., McDonald, J. C., Schueller, O. J. A. & Whitesides, G. M. Rapid prototyping of microfluidic systems in poly(dimethylsiloxane). *Analyt. Chem.* **70**, 4974–4984 (1998).
- Quake, S. R. & Scherer, A. From micro- to nanofabrication with soft materials. *Science* **290**, 1536–1540 (2000).
- McDonald, J. C. *et al.* Fabrication of microfluidic systems in poly(dimethylsiloxane). *Electrophoresis* **21**, 27–40 (2000).
- Thorsen, T., Maerkl, S. J. & Quake, S. R. Microfluidic large-scale integration. *Science* **298**, 580–584 (2002).
- Beebe, D. J. *et al.* Microfluidic tectonics: a comprehensive construction platform for microfluidic systems. *Proc. Natl Acad. Sci. USA* **97**, 13488–13493 (2000).
- Fiorini, G. S., Lorenz, R. M., Kuo, J. S. & Chiu, D. T. Rapid prototyping of thermoset polyester microfluidic devices. *Anal. Chem.* **76**, 4697–4704 (2004).
- Sudarsan, A. P., Wang, J. & Ugaz, V. M. Thermoplastic elastomer gels: an advanced substrate for microfluidic chemical analysis systems. *Anal. Chem.* **77**, 5167–5173 (2005).
- Rolland, J. P., Van Dam, R. M., Schorzman, D. A., Quake, S. R. & DeSimone, J. M. Solvent resistant photocurable 'liquid teflon for microfluidic device fabrication'. *J. Am. Chem. Soc.* **126**, 2322–2323 (2004).
- Xia, Y. N. *et al.* Complex optical surfaces formed by replica molding against elastomeric masters. *Science* **273**, 347–349 (1996).
- Gambin, Y., Legrand, O. & Quake, S. R. Microfabricated rubber microscope using soft solid immersion lenses. *Appl. Phys. Lett.* **88**, 174102 (2006).
- Introducing the Agilent Photonic Switching Platform [online] <http://physics.pdx.edu/~larosaa/Agilent_All_Optical_Network.htm>
- Campbell, K. *et al.* A microfluidic 2×2 optical switch. *Appl. Phys. Lett.* **85**, 6119–6121 (2004).
- Pang, L., Levy, U., Campbell, K., Groisman, A. & Fainman, Y. Set of two orthogonal adaptive cylindrical lenses in a monolith elastomer device. *Opt. Express* **13**, 9003–9013 (2005).
- Zhang, D. Y., Justis, N. & Lo, Y. H. Fluidic adaptive lens of transformable lens type. *Appl. Phys. Lett.* **84**, 4194–4196 (2004).
- Grillet, C. *et al.* Compact tunable microfluidic interferometer. *Opt. Express* **12**, 5440–5447 (2004).
- Heng, X. *et al.* Optofluidic microscopy — a method for implementing a high resolution optical microscope on a chip. *Lab Chip* (submitted).
- Domachuk, P., Littler, I. C. M., Cronin-Golomb, M. & Eggleton, B. J. Compact resonant integrated microfluidic refractometer. *Appl. Phys. Lett.* **88**, 093513 (2006).
- Liang, W., Huang, Y. Y., Xu, Y., Lee, R. K. & Yariv, A. *Appl. Phys. Lett.* **86**, 151122 (2005).
- Levy, U., Campbell, K., Groisman, A., Mookherjee, S. & Fainman, Y. On-chip microfluidic tuning of an optical microring resonator. *Appl. Phys. Lett.* **88**, 11107–11109 (2006).
- Homola, J., Yee, S. S. & Gauglitz, G. Surface plasmon resonance sensors: review. *Sens. Actuators B Chem.* **54**, 3–15 (1999).
- Schuck, P. Use of surface plasmon resonance to probe the equilibrium and dynamic aspects of interactions between biological macromolecules. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 541–566 (1997).
- Vo-Dinh, T. Surface-enhanced Raman spectroscopy using metallic nanostructures. *Trends Anal. Chem.* **17**, 557–582 (1998).
- Armani, A. M., Armani, D. K., Min, B., Vahala, K. J. & Spillane, S. M. Ultra-high-Q microcavity operation in H₂O and D₂O. *Appl. Phys. Lett.* **87**, 151118 (2005).
- Levene, M. J. *et al.* Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299**, 682–686 (2003).
- Wang, X., Wilson, D., Muller, R., Maker, P. & Psaltis, D. Liquid-crystal blazed-grating beam deflector. *Appl. Opt.* **39**, 6545–6555 (2000).
- Mach, P. *et al.* Tunable microfluidic optical fiber. *Appl. Phys. Lett.* **80**, 4294–4296 (2002).
- Li, Z. Y., Zhang, Z. Y., Emery, T., Scherer, A. & Psaltis, D. Single mode optofluidic distributed feedback dye laser. *Opt. Express* **14**, 696–701 (2006).
- Galas, J. C., Torres, J., Belotti, M., Kou, Q. & Chen, Y. Microfluidic tunable dye laser with integrated mixer and ring resonator. *Appl. Phys. Lett.* **86**, 264101 (2005).
- Vezenov, D. V. *et al.* A low-threshold, high-efficiency microfluidic waveguide laser. *J. Am. Chem. Soc.* **127**, 8952–8953 (2005).
- Helbo, B., Kragh, S., Kjeldsen, B. G., Reimers, J. L. & Kristensen, A. Investigation of the dye concentration influence on the lasing wavelength and threshold for a microfluidic dye laser. *Sens. Actuators A Phys.* **111**, 21–25 (2004).
- Joannopoulos, J. D., Meade, R. D. & Winn, J. N. *Photonic Crystals* (Princeton Univ. Press, Princeton, New Jersey, 1995).
- Domachuk, P., Nguyen, H. C., Eggleton, B. J., Straub, M. & Gu, M. Microfluidic tunable photonic band-gap device. *Appl. Phys. Lett.* **84**, 1838–1840 (2004).
- Erickson, D., Rockwood, T., Emery, T., Scherer, A. & Psaltis, D. Nanofluidic tuning of photonic crystal circuits. *Opt. Lett.* **31**, 59–61 (2006).
- Maune, B. *et al.* Liquid-crystal electric tuning of a photonic crystal laser. *Appl. Phys. Lett.* **85**, 360–362 (2004).
- Painter, O. *et al.* Two-dimensional photonic band-gap defect mode laser. *Science* **284**, 1819–1821 (1999).
- Dowling, J. P., Scalora, M., Bloemer, M. J. & Bowden, C. M. The photonic band edge laser: a new approach to gain enhancement. *J. Appl. Phys.* **75**, 1896–1899 (1994).
- Shinn, A. & Robertson, W. M. Surface plasmon-like sensor based on surface electromagnetic waves in a photonic band-gap material. *Sens. Actuators B Chem.* **105**, 360–364 (2005).
- Mookherjee, S. & Yariv, A. Coupled-resonator optical waveguides. *IEEE J. Select. Topics Quantum Electron.* **8**, 448–456 (2002).
- Knight, J. C., Broeng, J., Birks, T. A. & Russell, P. S. J. Photonic band gap guidance in optical fibers. *Science* **282**, 1476–1478 (1998).
- Cattaneo, F. *et al.* Digitally tunable microfluidic optical fiber devices. *J. Microelectromech. Syst.* **12**, 907–912 (2003).
- Xu, F. *et al.* Fabrication, modeling, and characterization of form-birefringent nanostructures. *Opt. Lett.* **20**, 2457–2459 (1995).
- Wolfe, D. B. *et al.* Dynamic control of liquid-core/liquid-cladding optical waveguides. *Proc. Natl Acad. Sci. USA* **101**, 12434–12438 (2004).
- Vezenov, D. V., Mayers, B. T., Wolfe, D. B. & Whitesides, G. M. Integrated fluorescent light source for optofluidic applications. *Appl. Phys. Lett.* **86**, 041104 (2005).
- Mugele, F. & Baret, J. C. Electrowetting: from basics to applications. *J. Phys. Condens. Matter* **17**, R705–R774 (2005).
- Hayes, R. A. & Feenstra, B. J. Video-speed electronic paper based on electrowetting. *Nature* **425**, 383–385 (2003).
- Groisman, A., Enzelberger, M. & Quake, S. R. Microfluidic memory and control devices. *Science* **300**, 955–958 (2003).
- Wolfe, D. B. *et al.* Diffusion-controlled optical elements for optofluidics. *Appl. Phys. Lett.* **87**, 181105 (2005).
- van der Hulst, H. C. *Light Scattering by Small Particles* (Dover Publications, New York, 1957).
- Domachuk, P. *et al.* Application of optical trapping to beam manipulation in optofluidics. *Opt. Express* **13**, 7265–7275 (2005).
- Brody, J. P. & Quake, S. R. A self-assembled microlensing rotational probe. *Appl. Phys. Lett.* **74**, 144–146 (1999).
- Trindade, T., O'Brien, P. & Pickett, N. L. *Chem. Mater.* **13**, 3843–3858 (2001).
- Lu, Y., Liu, G. L., Kim, J., Mejia, Y. X. & Lee, L. P. Nanophotonic crescent moon structures with sharp edge for ultrasensitive biomolecular detection by local electromagnetic field enhancement effect. *Nano Lett.* **5**, 119–124 (2005).
- Aizpurua, J. *et al.* Optical properties of gold nanorings. *Phys. Rev. Lett.* **90**, 057401 (2003).
- Neuman, K. C. & Block, S. M. Optical trapping. *Rev. Sci. Instrum.* **75**, 2787–2809 (2004).
- Grier, D. G. A revolution in optical manipulation. *Nature* **424**, 810–816 (2003).
- Chiou, P. Y., Ohta, A. T. & Wu, M. C. Massively parallel manipulation of single cells and microparticles using optical images. *Nature* **436**, 370–372 (2005).
- Adleman, J. R., Eggert, H. A., Buse, K. & Psaltis, D. Holographic grating formation in a colloidal suspension of silver nanoparticles. *Opt. Lett.* **31**, 447–449 (2006).
- Liu, G. L., Kim, J., Lu, Y. & Lee, L. P. Optofluidic control using photothermal nanoparticles. *Nature Mater.* **5**, 27–32 (2006).
- Whitesides, G. *et al.* Dynamic control of liquid-core/liquid-cladding optical waveguides. *Proc. Natl Acad. Sci. USA* **34**, 12434–12438 (2004).
- Erickson, D. Spectrographic microfluidic memory. *Proc. ICMM June 13–15 2005, Canada*.

Acknowledgements This work is funded by the Defense Advanced Research Projects Agency (DARPA) Center for Optofluidic Integration, USA. We thank J. Adelman, X. Heng, Y. Fainman and D. Erickson for numerous discussions and their assistance.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence should be addressed to D.P. (psaltis@sunoptics.caltech.edu).

Future lab-on-a-chip technologies for interrogating individual molecules

Harold Craighead¹

Advances in technology have allowed chemical sampling with high spatial resolution and the manipulation and measurement of individual molecules. Adaptation of these approaches to lab-on-a-chip formats is providing a new class of research tools for the investigation of biochemistry and life processes.

A range of powerful technologies exists for observing and identifying individual molecules. Well-established forms of microscopy such as electron microscopy and optical fluorescence microscopy have allowed us to image and, with the appropriate labels, identify the chemical nature of individual molecules. More recently, scanning probe microscopes have expanded the possibilities for observing and interacting directly with individual molecules. The fine probe of a scanning system can localize electrical measurements to a selected molecule, or measure the mechanical properties of a single biopolymer. Modern material-processing technologies, analogous to those that have been so successful in converting electronics to a 'chip-based' technology, are being explored for possible chemical or biological research lab-on-a-chip approaches¹.

Interest in scaled-down analytical processes, combined with advances in microfluidics, is motivating various chip-based methods in which analyses can be carried out more rapidly and at lower cost via small-scale systems than with current laboratory bench-scale methods. The new research approaches discussed below are motivated by the possibility of observing new phenomena or obtaining more detailed information from biologically active systems. For this, a class of research systems is evolving that uses a range of new physical and chemical approaches to biomolecular analysis. The complexity of life processes and the richness of molecular biology provide fertile ground for research with these new lab-on-a-chip approaches. The development of these new chip-based technologies is changing the nature of the questions that we can ask and for which we can seek experimental answers at the molecular level.

The developing chips are formed using technologies for inorganic-device processing combined with synthetic chemistry and biochemistry. The chips are beginning to integrate electrical, optical and physical measurements with fluid handling to create a new class of functional chip-based systems. The systems provide spatial localization that is relevant to observing, for example, the function of a single active enzyme, the activity of a single receptor on a cell surface, or the release of molecules from a single vesicle exocytotic event in an immune-system cell. A long-term result of this research could be a new class of fluid-handling chip systems engineered to use or analyse individual molecules. This could form the basis for ultra-sensitive sensors and medical diagnostic systems.

Molecular imaging and probes

The scanning probe instruments, such as the atomic force microscope and the scanning tunnelling microscope, have allowed the execution of a class of experiments in which individual molecules can be manipulated and probed. They are providing insights into biomolecular systems previously addressed only by indirect investigations. Scanning tunnelling microscopy can image molecules on surfaces and directly

observe their conformation and structure^{2,3}. Atomic force microscopy (AFM) can image molecules in a natural hydrated condition and capture images while the molecules are functioning⁴⁻⁶. The AFM has also been used to observe the folding and unfolding of individual protein molecules⁷⁻⁹. Observing the function of active molecules in a condition as close as possible to *in vivo* circumstances can provide new insights into how living systems function at the molecular level.

Ashkin was the first to use gradient forces from optical beams to trap particles and exert controlled forces on small beads, viruses and bacteria^{10,11}. The technique, which came to be known as 'optical tweezers', proved to be a powerful tool for applying controlled forces to individual molecules¹²⁻¹⁵. Typical experiments involve chemically binding a molecule of interest to a bead and chemically linking the other end of the molecule to a surface or another bead. The optical tweezers can be used in fluids and, as a result, can be used for biomolecules in a natural hydrated and functioning condition. This permits force-distance measurements as a biopolymer is unfolded, or the investigation of the breaking of bonds in specific reactions. This technique is now widely used in biophysical studies of single molecules. It has done a great deal to move researchers towards thinking of single molecules as accessible objects that can be selected and studied for their individual properties. It has also helped motivate chip-based approaches for individual-molecule analysis.

Patch-clamp technology, another form of localized electrical probe, revolutionized our ability to study active ion channels in cell membranes¹⁶. The technology uses a glass pipette drawn down to form a micrometre-scale aperture that can make a tight, electrically insulating seal to a cell membrane. With the electrical current conducted and measured through the pipette, electrical charge is constrained to conducting paths inside the sealed region of the pipette tip. This was a revolutionary technique that allowed the observation of single-molecule ion channels in living cell membranes.

Although powerful and productively used in experimental research, the single probe methods use tedious manipulation of the individual probes and do not take advantage of the integration, parallelism and automation available from a chip-based system. They do, however, set the stage for migration of these approaches to chip-based methods, and possibly further integration into research systems. Researchers are beginning to develop and use engineered devices that are often fabricated in large arrays for accessing individual molecules.

Developing technologies for lab-on-a-chip integration

Technologies exist that unite lithographic approaches for processing hard materials with soft material processing, fluidics and biochemical patterning. These technologies, which are related to those used in commercial device manufacturing, reliably produce systems with feature

¹Applied and Engineering Physics, 205 Clark Hall, Cornell University, Ithaca, New York 14853, USA.

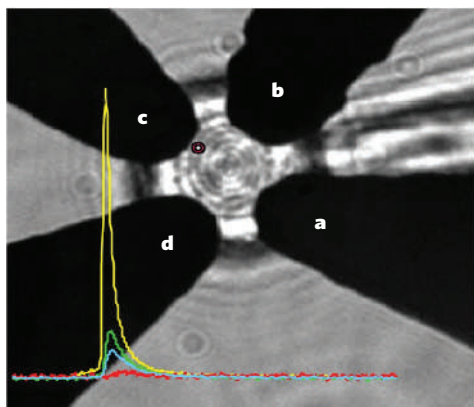


Figure 1 | Optical micrograph of a 4-electrode electrochemical detector array on which a chromaffin cell has been placed. The superimposed trace is the measured current as a function of time at electrodes **a–d** (**a**, red; **b**, green; **c**, yellow; **d**, blue). The small circle drawn near electrode **c** indicates the location of the fusion pore opening, calculated from the relative current collected from the four electrodes⁴¹.

sizes of less than 100 nm^{17–22}. They are also used to create complex and massively parallel electronic devices, creating billions of essentially identical devices in a chip format. Optical sources and detectors have also been miniaturized and made parallel, and are available in consumer products such as video cameras and displays. This is a powerful and well-developed technology that can be exploited in the research arena today, and possibly in the future in a new class of highly-functional chip-like devices for biochemical analysis.

At this stage of research, many different materials are being explored and used. Polymers are attractive for many microfluidic uses because they are easily manufactured by embossing or moulding²³. They can also be bonded to other surfaces and are therefore often used to form a fluid channel on another surface, such as silicon, on which an electronic device has been formed. Single-molecule approaches using electrical or optical detection may create other material requirements with regard to electrical or optical properties. Semiconductors and metals are obviously necessary components of electrical detection schemes. Semiconductor nanowires and carbon nanotubes, for example, are being studied as sensor components^{24–26}. Mechanical devices are also being explored and integrated in fluid systems^{27–30}. In addition, porous media for sample concentration and filtration can be formed in fluid channels³¹. The situation is nothing like that of silicon microelectronics technology, in which the technology is mature and material technologies are monolithic. However, some common materials and processing techniques are evolving, often on the basis of silicon compounds and polymers, that suggest the possibility for greater integration of fluid handling with optical, electrical and mechanical devices.

Electrical chips with high spatial resolution

The development of the planar patch clamp is an example of the conversion of a very successful probe technique to a chip-type device^{32–39}. In this approach, rather than using the narrow aperture of a drawn glass pipette, an electrical probe is formed by an etched opening, of the order of a micrometre, that can be readily fabricated in a thin membrane formed on a planar surface. The macroscopic electrodes are then placed on opposite sides of the aperture, with all current forced to traverse the narrow aperture. The entity of interest, typically an ion channel in a lipid layer spanning the aperture, is thus electrically isolated. Schmidt, Mayer and Vogel³² demonstrated such a device in a silicon nitride membrane supported on an etched silicon wafer, used to observe the activity of a single voltage-gated ion channel of the peptide alamethicin integrated in a lipid bilayer. The use of a fabrication approach with a lithographically defined aperture formed a reliable aperture size, and opened up the possibility of making arrays of many apertures for localized electrochemical measurement. Although they do not replace the traditional

moveable probe technique, microfabricated planar patch-clamp-like devices such as this are being used. Their planar geometry allows them to be engineered into sensor configurations.

Electronic-chip-fabrication technology can be used to create arrays of electrodes for probing discrete electrical events in living systems. Lindau's group, for example, described an electrode array for spatio-temporal resolution of single exocytotic events in living cells^{40,41}. Using a patterned metallic electrode array, on which the cells were placed, electrochemical signals were recorded from the oxidation of catecholamines released by the cell during events in which a vesicle contained in the cell fused with the cell membrane, releasing its contents. By measuring the relative strengths of the current signals from each electrode, the researchers could calculate the position of each individual exocytotic event as well as the total quantity of ions released and the time taken by the event. This level of quantitative detail is difficult to obtain by other methods and, again, the format allows for scaling the number of groups of identical electrodes for measurements, in principle, on large arrays of cells. This is of interest for a number of basic cell-biology studies, and is also of potential utility in obtaining more detailed information about the effects of potential drugs on cellular function.

Control of the structure and design of the arrays also enables integration with complementary measurement approaches. With the same types of electrode array (Fig. 1), Lindau and colleagues were also able to record simultaneously fluorescence images of fusion-pore events in the cells. This brought to bear the full capability of chemically selective fluorescent dyes and optical imaging to verify the interpretation of the electrochemical imaging. This also indicates the value of this type of lab-on-a-chip in terms of engineering the planar devices for integration for multiple analytical processes.

Of course, the capabilities of integrated electronics extend well beyond those of passive electrodes; researchers can incorporate active electronic devices to allow more complex functions, such as signal amplification, to be carried out. Many groups have used microfabricated electrode arrays for extracellular recording of action potentials from

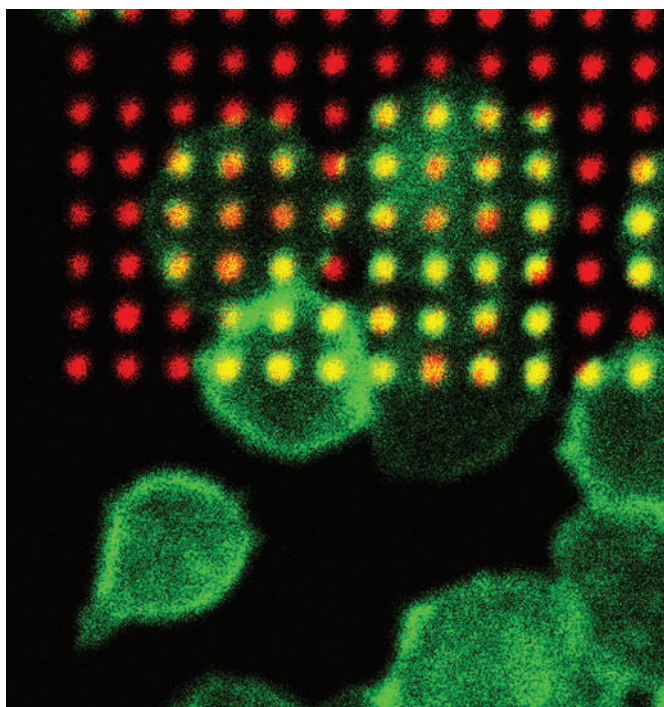


Figure 2 | Fluorescence micrograph of micrometre-size patterned lipid bilayers containing specific ligands used to cluster mast-cell receptors. Clustered receptors (green) stimulate transmembrane signalling events, and the patterns (red) allow visualization of spatial regulation of reorganizing cellular components. Co-concentrated red and green fluorescence appears yellow^{45,46}. (Image courtesy of B. Baird, Cornell University, USA.)

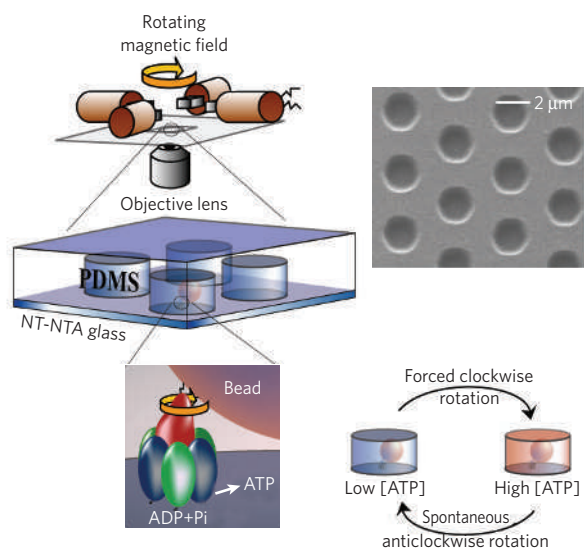


Figure 3 | Schematic of F_1 ATPase rotary motor enzyme to which a magnetic bead is attached. The bead can be used to drive or observe the motion of the motor. Each enzyme is isolated in a reaction chamber of radius $\sim 1 \mu\text{m}$ (shown schematically and in an electron micrograph) that contains the reaction products⁴⁷. (Image courtesy of H. Noji, University of Tokyo, Japan.)

neurons cultured on the electrode arrays. The Frommertz group has used arrays of electrodes connected to on-chip arrays of transistors to amplify voltage signals near their source in order to improve the signal-to-noise ratio⁴². With the provision of active electronics to the chip, additional signal processing and logic functions could be incorporated into the device.

In addition to the integration of electrical function with optical analysis, the chip-oriented approach can be coupled with biochemical patterning to interface more effectively to biological systems. For example, James *et al.* used methods for patterning proteins in registry with electrode arrays^{43,44}. The patterned proteins guided the development of neuronal-cell growth on the electrode surface that was used to record signals from selected points in an organized cell culture. The ability to use patterned cell-growth factors, chemotactic agents or selective binding factors, such as antibodies, provides a method for interfacing the powerful analytical capabilities engineered into the inorganic devices with organized cellular systems.

As noted in the Review on cell biology in this issue (page 403), patterning of active biomaterials on surfaces can be used for cell-based lab-on-a-chip assays. However, patterning at subcellular dimensions makes it possible to orient cells with respect to chip elements. The patterns themselves can be used to investigate cell responses to localized stimuli and to observe molecular-scale responses. For example, Orth *et al.* used a high-resolution method of patterning to create patterned lipid bilayers containing antigens to mammalian immune-system cells⁴⁵. The cell-surface receptors clustered in the same pattern and resulted in cell activation. Wu *et al.* used this approach to investigate targeting in transmembrane signalling events⁴⁶. Using fluorescence microscopy with exogenous or genetically encoded fluorescent probes, the stimulated redistribution membrane, cytoskeletal and cytoplasmic components were tracked (Fig. 2). This allowed the testing of hypotheses about the formation of so-called lipid rafts, and demonstrated decoupling of the inner and outer leaflets of the cell membrane. Similar to the previous examples, the ability to create large numbers of essentially identical chemical stimuli allowed the system to be replicated and presented to numerous cells. The use of the patterned biochemical chip in conjunction with fluorescence microscopy provided an unambiguous visualization of the cellular response to defined stimuli and the capability to quantify it.

Planar devices for single-molecule observation

Simple chips with nanoscale features can be used as windows onto the activity of single molecules, extending the capability for observation of life processes to the molecular scale. The Noji group created an array of individual femtolitre chemical chambers on a chip for studies of the activity of individual F_1 ATPase enzyme molecules bound to a surface^{47,48}. In this work, the group formed small, optically transparent polymeric chambers to isolate the chemistry associated with an individual enzyme (Fig. 3). The ATPase enzyme is a rotary motor, and by attaching a magnetic bead to the rotating part of the molecule, the rotation of the molecule could be observed optically. The torque on the magnetic bead could be controlled by application of a rotating magnetic field. The researchers were able to count the rotations of the motor and calculate the efficiency as the enzyme motor consumed its ATP fuel. Because the volume of solution containing ATP was very small, the individual enzyme's activity in depleting the available energy source could be observed in reasonable experimental timescales. In a demonstration of incredible control, the individual motors could be driven in the opposite direction by the rotating magnetic field to synthesize countable ATP molecules.

This experiment provides unique insights into the activity of functioning biomolecules, directly addressing questions of kinetics, rates and efficiencies. This is an example of a potentially revolutionary class of chip-based devices in which the observation and control of large numbers of individual molecules are used as detectors of identifiable individual chemical events or components of a chemical synthesis system.

Active enzymes can be optically isolated for measurement by techniques that allow the observation of individual reaction events at high concentrations. The previously discussed approach used physical confinement of reactants to limit the reaction to a small volume. With optical confinement to a small volume surrounding an active enzyme, the observation of optically stimulated or interrogated processes can be localized to observe chemical activity specific to the isolated enzyme. Levene *et al.* described an array of subwavelength-diameter metallic apertures (Fig. 4) that confine light to dimensions well below the diffraction limit in three dimensions⁴⁹. Light incident on this metallic structure, which can be considered as a metallic waveguide operated at a frequency below the cutoff frequency, is rapidly attenuated in the direction of propagation. Because it has no propagating waveguide modes, it is termed a zero-mode waveguide. With a diameter of a few tens of nanometres, this attenuation length becomes less than the film thickness. The transverse dimensions of the light are, of course, limited by the dimensions of the aperture. Other means of optical confinement include the use of metallic tips to concentrate optical excitation⁵⁰.

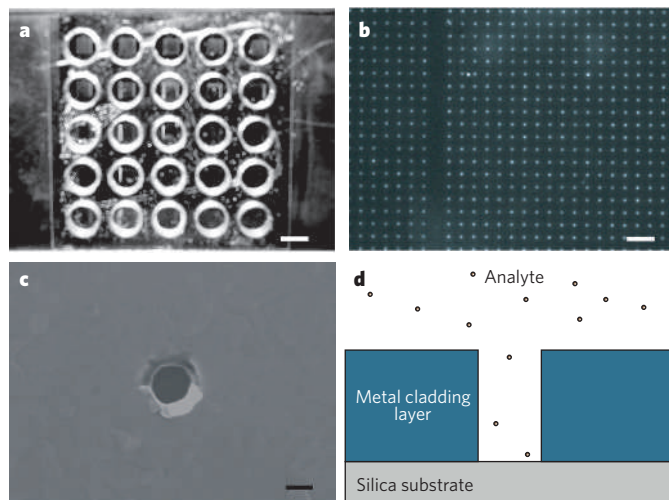


Figure 4 | Arrays of metallic apertures used for optical observation of individual molecules. Array of polymer wells containing arrays of zero-mode waveguides (a) shown in electron micrograph (b) and in higher magnification (c). Also shown (d) is a schematic cross-section of the metallic aperture with analyte solution on the metal side⁴⁹.

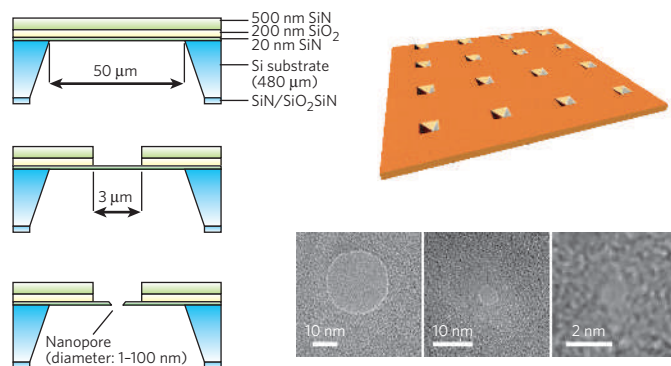


Figure 5 | Nanopores used for DNA translocation studies. Left: cross-sectional schematic of an engineered nanopore in silicon nitride showing three phases of the pore fabrication. Top right: AFM image of an array of engineered pores formed on a chip. Bottom right: transmission electron micrograph of three different pores. (Images courtesy of C. Dekker⁸³, University of Delft, The Netherlands, and the American Chemical Society.)

The net effect is an optical excitation volume of the order of zeptolitres (10^{-21}), significantly smaller than could, for example, be obtained by total internal reflection illumination, which can confine light to the evanescent field in one direction only. If the light is used to stimulate fluorescence in molecules in a solution covering the zero-mode waveguide at a concentration, C , the average number of molecules, N , in the optical excitation volume, V , is $N=CV$. With an excitation volume of 10^{-21} litres, $N=1$ at a concentration of 10^{21} litres⁻¹, or 1.7 mM. The behaviour of a single optically labelled entity can therefore be observed at high concentrations. If the fluorescent molecule is freely diffusing in the liquid, then the residence time in the optical excitation volume is determined by the diffusion coefficient of the molecule. By observing this time for individual molecules, the concentration of entities with different diffusion coefficients can be determined. This can be used, for example, to measure the degree of oligomerization of a biopolymer⁵¹. The differences in temporal fluorescent behaviour are even more pronounced in cases in which a species is permanently bound, as has been observed for the enzymatic synthesis of double-stranded DNA by DNA polymerase⁴⁹. In this case, non-bound fluorescent species are only transiently fluorescent during their diffusive motion through the excitation region. Bound species are fluorescent for times limited by fluorescent bleaching or other processes for termination of the fluorescence. With an array, a large number of independent observation volumes can be used for the observation of enzymatic or other optically distinguishable chemical events^{49–53}.

Devices consisting of nanoscale pores in an electrically insulating layer are being pursued as probes of molecular structure^{54–59}. In these devices, ion current passing through the narrow pore is modulated, as it passes through the pore, by the presence or nature of a large biopolymer molecule such as DNA. Naturally occurring membrane proteins such as α -haemolysin have well-defined pore dimensions, of the order of a nanometre. Engineered pores, fabricated by various forms of etching, can form similar apertures in an insulating membrane. Figure 5 shows a schematic and electron micrograph, from the Dekker group, showing apertures a few nanometres in diameter formed in a silicon nitride membrane⁵⁶. When DNA in solution is electrically driven through such a pore, the presence of the DNA influences the passage of other mobile ions in solution through the narrow pore, and the transit of the DNA can be detected by changes in the current measured through the pore. In the simplest case, if the electrically driven speed of the molecule is known, the length of an extended DNA molecule can be inferred from the transit time. Confirmation of the molecule is observed as discrete differences in the current levels corresponding to the condition of a single strand — or a double, or triple, and so on — of a folded DNA molecule passing through the pore. Because of the strong motivation for rapid genetic sequencing or analysis, efforts are under way to extract

the identity of DNA bases by resolvable electrical differences as they pass through the pore, but the resolution requirements, signal-to-noise ratio in the current signal, and the subtle differences in electrical signatures of different bases make this a difficult task. The nanopores, however, are simple but powerful biophysical probes of molecular confirmation, and a means of observing and controlling the forces on biopolymers in fluid.

Single-molecule optical analysis in flowing systems

Optical techniques are being developed for observing, detecting, analysing and quantifying single molecules. Labelling a molecule of interest with a bright fluorescent dye molecule or luminescent semiconductor particle allows selected molecules to be located and tracked by optical microscopy. Combining chemically-specific binding with the label makes it possible to identify a specific molecular species. The general approach of selective fluorescent labelling is widely used in imaging and immunoassays, and, as a result, labelling chemistry is well developed. Adapting the fluorescent approaches to analysis requires integration with fluidics and dealing with the noise limits inherent in single-molecule detection. In addition to the chemical identification imparted by specific binding chemistry, single-molecule methods can also be used to identify other features of a molecule, such as the diffusion rate, electrophoretic mobility, or rate of passage through a mechanical constriction or pore. In appropriate systems, the single-molecule approach can provide unique possibilities in quantification and dealing with limited samples, such as one from a single cell. Further to identification of chemical species or character, a significant application of single-molecule analysis is geared towards extracting genetic information from nucleic acids. A set of techniques is being developed for analysing individual molecules, with the goal of more rapid genetic or RNA expression analysis.

Optical approaches allow molecules of interest to be highlighted by labelling, but to make single-molecule approaches practical this must be coupled to microfluidic systems to efficiently deliver molecules to the optical analytical system^{60–70}. Confocal optical approaches provide a direct method with which to excite and detect fluorescence from a microfluidic system. A basic system is shown in the schematic of Fig. 6. Current research approaches tend to use large-scale microscopes, lasers

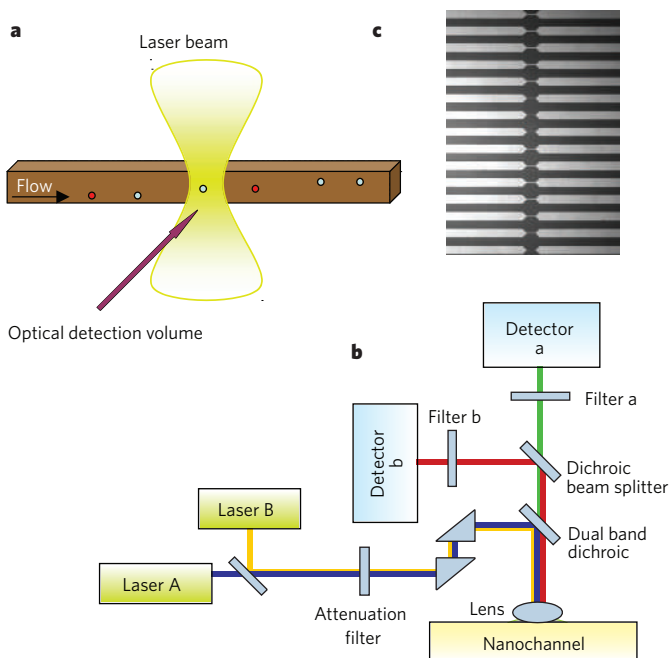


Figure 6 | Fluidic channel system for single-molecule optical measurements. a, Schematic of a fluid channel with optical excitation and detection volume. b, Schematic of optical set up. c, Image of 15 parallel fluid channels (pale grey) with constricted analysis region (centre)⁶¹.

and detectors because of the availability of these systems. It should be clear, however, that the optical systems could also be miniaturized into more integrated and functional chips.

One unique single-molecule approach is to count individual molecules as they pass an interrogation region. With spectrally identifiable labels, the exact concentrations of different species could be obtained by directly counting the numbers of each molecular species. As signal-to-noise ratio is important, it is desirable to limit the size of the excitation volume to reduce background from scattering or intrinsic fluorescence of unlabelled species in the excitation volume. This motivates the use of confinement of the excitation by limiting either the optical illumination size or the physical size of the channel. Decreasing the optical excitation below the channel width reduces the detection efficiency and makes it possible that not all molecules in the solution will be detected. Reducing the physical size of the channel improves the optical signal-to-noise ratio, and increases the interaction with the channel surfaces. The optimum in the situation is dictated by the nature of the analysis to be performed, including the concentration range of the molecules, the volume of sample and the strength of surface interactions. It seems likely that devices will be designed for specific applications.

Single-molecule approaches make it possible to detect single binding events, in addition to simply counting differentially labelled species. For example, individual binding events can be detected by simultaneously detecting the presence of fluorescence from two differently labelled molecules. Labels can be formed by pairs of molecules in which excitation energy is transferred from the donor member of the pair to the acceptor of fluorescent excitation in a fluorescence-resonance-energy-transfer or Förster-resonance-energy-transfer (FRET) process. In this case, the emission will be observed only when the binding takes place, bringing donor and acceptor into close proximity. In a microfluidic system in which the molecule of interest has a charge and an electric field is used to drive the molecule, a measure of the electrophoretic mobility can be determined from the transit time through a channel of known length⁶⁴. Changes in mobility could indicate the difference between an unbound labelled capture molecule and a labelled capture molecule bound to a protein. The combination of microfluidics and electrically or pressure driven fluid systems creates integrated design possibilities. Parallelism from many fluid and optical channels could be designed with modern optoelectronic technology. Planar waveguide excitation, out of plane diffractive optics, laser arrays and detector arrays are all readily available to be adapted for miniaturized systems.

In addition to counting and determining the chemical identity of molecules from specific binding chemistry, the conformation and physical properties of individual molecules can be investigated in nanofluidic systems^{71–80}. This can be considered to be the chip-based analogue of the probe and optical-tweezer techniques discussed above. A number of studies have been done on the sorting of DNA molecules by size in nanostructures. When associated with restriction digests, the sizes of the DNA fragments contain information on the sequence information in the original longer DNA molecules. Rapid measurement of DNA fragment size was demonstrated in a series of nanostructures with dimension smaller than the radius of gyration of the DNA molecule in free solution. In the so-called entropic traps^{75,78}, the rate at which a molecule could enter the constriction was length-dependent, and this dependence was used to separate bands of DNA in electrophoretically driven microfluidic channels with nanoscale constrictions. The separation was much more rapid than that achievable with conventional gel-based techniques.

Similar entropic forces in nanochannels have been used to control forces on individual molecules and to elongate them for study^{71–80}. When DNA is driven into a fluid channel with transverse dimensions much less than the radius of gyration, or, even more critically, when the dimensions become comparable to the persistence length of the polymer, the confinement exerts forces on the molecule that push it from the constrained region⁷³ (Fig. 7). The effect of the entropic forces can be used to measure effectively the length of the molecule by the rate of passage into the constrained region⁷⁶. Control of the molecular conformation also allows analysis of DNA in the context of the overall chromosome.

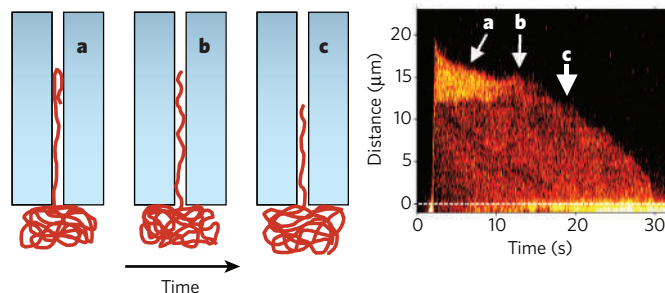


Figure 7 | Schematic and images of DNA retraction from a nanochannel. Diagram (left) showing three different time stages (a–c) of the entropically driven motion of a single DNA molecule at the interface between a microchannel and a nanochannel. After being electrophoretically driven into the channel as a ‘hair pin’ loop, the molecule unfolds as it is entropically pulled from the nanochannel by forces acting only on the long end. The image on the right shows stacked fluorescence images of a molecule as a function of time, demonstrating the recoiling and straightening process. The looped (a) and then the straightened (b, c) condition of the molecule are indicated. Once straightened, the molecule can be manipulated and positioned in the channel as an unfolded linear molecule⁷³.

Towards this end, Austin’s group demonstrated the activity of restriction enzymes, cutting DNA at prescribed sequence locations, in an extended single molecule in a nanochannel⁷⁷.

Because of interest in their underlying sequence, DNA and RNA represent important targets for single-molecule studies, but the same concepts and physical methods of characterizing biomolecules are being extended to proteins and other compounds. In cases in which chemical differences accompany the physical differences, the situation is more complex, complicating the interpretation of phenomena for analysis. The approach for analysis, however, combining microfluidics, nanoscale structures and optical detection, creates a range of opportunities for molecular analysis.

Outlook for highly functional integrated systems

The current types of research-chip device comprise fairly simple structures, as we have seen above, such as two-dimensional arrays of apertures or reaction chambers to isolate individual molecules for study. In these simple initial devices the advantages of massively parallel replication of identical units with 10^5 , 10^6 or more individual reaction or observation areas can be seen. In fluid channel systems that permit one-dimensional arrays, the degree of parallelism could easily be $\sim 10^3$. It is clear that the scale and levels of integration are compatible with active optical and electronic devices in today’s integrated electronic and optoelectronic systems, and carrying out millions of parallel optical or electronic measurements is a realistic possibility. Because optical beams can traverse free space and fluid systems can be made transparent, the integration with external optical systems is possibly the most straightforward initial direction of integration. Electrical techniques require direct contact with the analyte and generally also require wires for interconnects, which creates more challenges for integration. The advances in this area are likely to be important for determining the breadth of utility of such lab-chip devices.

Opportunities for single-molecule analysis include the ability to characterize small volumes of complex mixtures such as one would obtain from a single cell. The ability to characterize the differences in populations of cells could be greatly advanced by lab-on-a-chip approaches that could rapidly quantify the levels of a set of proteins of interest in a selected cell at a particular time. As previously noted, the ability to handle cells and present controlled stimuli to cells is developing along with techniques that could help to analyse the function of these cells. In a similar way, rapid detection of the expression of RNA or the genetic make-up of individual cells could be accessed by such analytical systems. The capacity for sensitive and rapid analysis of the complex chemical content of body fluids could enable new medical diagnostic approaches, identifying the markers for disease at a stage at which treatment can be most effective.

The goal of substantially more rapid and inexpensive sequencing of entire genomes is now well established. The '\$1,000 genome' is now a stated target of the US National Institutes of Health⁸¹. A range of approaches to this goal is being considered, including single-molecule approaches. The combination of greater understanding of biopolymer physics and the ability to integrate fluidics is presenting new opportunities and research directions.

A valuable device would allow rapid, less invasive analysis of complex biological fluids for medical diagnosis and monitoring of therapies. A range of technologies is being directed towards this general goal. The tool box of lab-on-a-chip methods, and particularly single-molecule approaches, may, in the long term, enable the engineering of a number of highly functional devices for specific diagnostic targets.

Biological and medical applications are viewed as areas of high potential impact of single-molecule approaches. Perhaps a biological analogy could provide insights into how engineered systems with controllable enzymes could be used to 'manufacture' small amounts of a large number of compounds in combinatorial approaches. Large libraries of such synthesized compounds could then be studied by comparably miniaturized systems. In any case, the existence of natural biological systems that have the ability to sense remarkably small numbers of molecules in chemically noisy environments provides guidance on how to best engineer artificial systems to operate most effectively in environments in which the statistics of signalling with small numbers of molecules⁸² and time limits associated with molecular motion are the fundamental limits of operation. Living cells have evolved clearly viable approaches to sensing and reacting to their environment on the basis of the analysis of single-molecule events. ■

- Manz, A. *et al.* Miniaturized total chemical analysis systems — a novel concept for chemical sensing. *Sens. Actuators B* **1**, 244–248 (1990).
- Muller, D. J., Amrein, M. & Engle, A. Adsorption of biological molecules to a solid support for scanning probe microscopy. *J. Struct. Biol.* **119**, 172–188 (1997).
- Stipe, B. C. A variable temperature scanning tunneling microscope capable of single-molecule vibrational spectroscopy. *Rev. Sci. Instr.* **70**, 133–136 (1999).
- Hansma, H. & Hoch, J. Biomolecular imaging with the atomic force microscope. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 115–134 (1994).
- Bustamante, C. & Keller, D. Scanning force microscopy in biology. *Phys. Today* **48**, 32–38 (1995).
- Reif, M. *et al.* Single molecule force spectroscopy on polysaccharides by atomic force microscopy. *Science* **275**, 1295–1297 (1997).
- Reif, M. *et al.* Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **276**, 1109–1112 (1997).
- Carrión-Vázquez, M. *et al.* Mechanical and chemical unfolding of a single protein: a comparison. *Proc. Natl Acad. Sci. USA* **96**, 3694–3699 (1999).
- Fisher, T. E. Stretching single molecules into novel conformations using the atomic force microscope. *Nature Struct. Biol.* **7**, 719–724 (2000).
- Ashkin, A. *et al.* Observation of a single-beam gradient force optical trap for dielectric particles. *Opt. Lett.* **11**, 288–290 (1986).
- Ashkin, A. *et al.* Optical trapping and manipulation of single cells using infrared laser beams. *Nature* **330**, 769–771 (1987).
- Wang, M. D. *et al.* Stretching DNA with optical tweezers. *Biophys. J.* **72**, 1335–1346 (1997).
- Grier, D. G. A revolution in optical manipulation. *Nature* **424**, 810–816 (2003).
- Sisichka, A. *et al.* Compact microscope-based optical tweezers system for molecular manipulations. *Rev. Sci. Instr.* **74**, 4827–4831 (2003).
- Mehta, A. *et al.* Single-molecule biomechanics with optical methods. *Science* **283**, 1689–1695 (1999).
- Neher, E. & Sakmann, B. Single-channel currents recorded for membrane of denervated frog muscle fibres. *Nature* **260**, 799–802 (1976).
- Mijatovic, D., Eijkel, J. C. & van den Berg, A. Technologies for nanofluidic systems: top-down vs. bottom-up — a review. *Lab Chip* **5**, 492–500 (2005).
- Craighead, H. G. Nanoelectromechanical systems. *Science* **290**, 1532 (2000).
- Beebe, D. J., Mensing, G. A. & Walker, G. M. Physics and applications of microfluidics in biology. *Annu. Rev. Biomed. Eng.* **4**, 261–286 (2002).
- Quake, S. R. & Scherer, A. From micro- to nanofabrication with soft materials. *Science* **290**, 1536–1540 (2000).
- Schwarz, M. A. & Hauser, P. C. Recent developments in detection methods for microfabricated analytical devices. *Lab Chip* **1**, 1–6 (2001).
- Liu, J., Hansen, C. & Quake, S. R. Solving the "world to chip" interface problem with a microfluidic matrix. *Anal. Chem.* **75**, 4718–4723 (2003).
- Ng, J. M. K., Gitlin, I., Stroock, A. D. & Whitesides, G. M. Components for integrated poly(dimethylsiloxane) microfluidic systems. *Electrophoresis* **23**, 3461–3473 (2002).
- Zheng, G. *et al.* Multiplexed electrical detection of cancer markers with nanowire sensor arrays. *Nature Biotechnol.* **23**, 1294–1301 (2005).
- Patolsky, F. & Lieber, C. M. Nanowire nanosensors. *Mater. Today* **8**, 20–28 (2005).
- Rosenblatt, S. High performance electrolyte-gated carbon nanotube transistors. *Nano Lett.* **2**, 869–872 (2002).
- Sazonova, V. *et al.* A tunable carbon nanotube electromechanical oscillator. *Nature* **431**, 284–287 (2005).
- Burg, T. & Manalis, S. Suspended microchannel resonators for biomolecular detection. *Appl. Phys. Lett.* **83**, 2698–2701 (2003).
- Verbridge, S. *et al.* Suspended glass nanochannels coupled with microstructures for single molecule detection. *J. Appl. Phys.* **97**, 124317–124320 (2005).
- Ilic, B. *et al.* Enumeration of DNA molecules bound to a nanomechanical oscillator. *Nano Lett.* **5**, 925–929 (2005).
- Song, S., Singh, A. K. & Kirby, B. J. Electrophoretic concentration of proteins at laser-patterned porous membranes. *Anal. Chem.* **76**, 4589–4592 (2004).
- Schmidt, C., Mayer, M. & Vogel, H. A chip-based biosensor for the functional analysis of single ion channels. *Angew. Chem. Int. Ed.* **39**, 3137–3140 (2000).
- Suzuki, H. *et al.* Highly reproducible method of planar lipid bilayer reconstitution in polymethyl methacrylate microfluidic chip. *Langmuir* **22**, 1937–1942 (2006).
- Fertig, N. *et al.* Microstructured glass chip for ion-channel electrophysiology. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* **64**, 040901–1–040901–4 (2001).
- Fertig, N., Blick, R. H. & Behrends, J. C. Whole cell patch clamp recording performed on a planar glass chip: electrophysiology on a chip. *Biophys. J.* **82**, 3056–3062 (2002).
- Klemic, K. B. *et al.* Micromolded PDMS planar electrode allows patch-clamp electrical recordings from cells. *Biosens. Bioelectron.* **17**, 597–604 (2002).
- Groves, J. T., Ulman, N. & Boxer, S. G. Micropatterning fluid bilayers on solid supports. *Science* **275**, 651–653 (1997).
- Pantoja, R. *et al.* Bilayer reconstitution of voltage-dependent ion channels using a microfabricated silicon chip. *Biophys. J.* **81**, 2389–2394 (2001).
- Xu, L. *et al.* Ion-channel assay technologies: quo vadis? *Drug Discov. Today* **6**, 1278–1287 (2001).
- Dias, A. *et al.* An electrochemical detector array to study cell biology on the nanoscale. *Nanotechnology* **13**, 285–289 (2002).
- Hafez, I. *et al.* Electrochemical imaging of fusion pore openings by electrochemical detector arrays. *Proc. Natl Acad. Sci. USA* **102**, 13879–13884 (2005).
- Zech, G. & Frommertz, P. Noninvasive neuroelectronic interfacing with synaptically connected snail neurons immobilized on a semiconductor chip. *Proc. Natl Acad. Sci. USA* **98**, 10457–10462 (2001).
- James, C. D. *et al.* Extracellular recordings from patterned neuronal networks using planar microelectrode arrays. *IEEE Trans. Biomed. Eng.* **51**, 1640–1648 (2004).
- Oliva, A. A. *et al.* Patterning axonal guidance molecules using a novel strategy for microcontact printing. *Neurochem. Res.* **28**, 1639–1648 (2003).
- Orth, R. N. *et al.* Mast cell activation on patterned lipid bilayers of subcellular dimensions. *Langmuir* **19**, 1599–1605 (2003).
- Wu, M. *et al.* Visualization of plasma membrane compartmentalization with patterned lipid bilayers. *Proc. Natl Acad. Sci. USA* **101**, 13798–13803 (2004).
- Rondelez, Y. *et al.* Highly coupled ATP synthesis by F₁-ATPase single molecules. *Nature* **433**, 773–777 (2005).
- Rondelez, Y. *et al.* Microfabricated arrays of femtoliter chambers allow single molecule enzymology. *Nature Biotechnol.* **23**, 361–365 (2005).
- Levene, M. J. *et al.* Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299**, 682–686 (2003).
- Gerton, J. M., Wade, L. A., Lessard, G. A., Ma, Z. & Quake, S. R. Tip-enhanced fluorescence microscopy at 10 nanometer resolution. *Phys. Rev. Lett.* **93**, 180801 (2004).
- Samiee, K. T. *et al.* λ-Repressor oligomerization kinetics at high concentrations using fluorescence correlation spectroscopy in zero-mode waveguides. *Biophys. J.* **88**, 2145–2153 (2005).
- Samiee, K. *et al.* Zero mode waveguides for single molecule spectroscopy on lipid membranes. *Biophys. J.* **90**, 3288–3299 (2006).
- Rigneault, H. J. *et al.* Enhancement of single molecule fluorescence detection in subwavelength apertures. *Phys. Rev. Lett.* **95**, 117401–117404 (2005).
- Kasianowicz, J. *et al.* Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl Acad. Sci. USA* **93**, 13770–13773 (1996).
- Deamer, D. W. & Branton, D. Characterization of nucleic acids by nanopore analysis. *Acc. Chem. Res.* **35**, 817–825 (2002).
- Storm, A. J. *et al.* Fast DNA translocation through a solid-state nanopore. *Nano Lett.* **5**, 1193–1197 (2005).
- Heng, J. B. *et al.* Sizing DNA using a nanometer-diameter pore. *Biophys. J.* **87**, 2905–2911 (2004).
- Li, J. *et al.* DNA molecules and configurations in a solid state nanopore microscope. *Nature Mater.* **2**, 611–615 (2003).
- Mathe, J. *et al.* Nanopore unzipping of individual DNA hairpin molecules. *Biophys. J.* **87**, 3205–3212 (2004).
- Goodwin, P. M., Nolan, R. L. & Cai, H. Single-molecule spectroscopy for nucleic acid analysis: a new approach for disease detection and genomic analysis. *Curr. Pharm. Biotechnol.* **5**, 271–278 (2004).
- Stavis, S. M. Single molecule studies of quantum dot conjugates in a submicrometer fluidic channel. *Lab Chip* **5**, 337–343 (2005).
- Stavis, S. M. Detection and identification of nucleic acid engineered fluorescent labels in submicrometer fluidic channels. *Nanotechnology* **16**, S314–S323 (2005).
- Foquet, M. DNA fragment sizing by single-molecule detection in submicrometer-sized closed fluidic channels. *Anal. Chem.* **74**, 1415–1422 (2002).
- Stavis, S. M. Single-molecule mobility and spectral measurements in submicrometer fluidic channels. *J. Appl. Phys.* **98**, 044903–1–044903–5 (2005).
- Campbell, L. C. *et al.* Electrophoretic manipulation of single DNA molecules in nanofabricated capillaries. *Lab Chip* **4**, 225–229 (2004).
- Mogensen, K. B., El-Ali, J., Wolff, A. & Kutter, J. P. Integration of polymer waveguides for optical detection in microfabricated chemical analysis systems. *Appl. Opt.* **42**, 4072–4079 (2003).
- Foquet, M. E. *et al.* Fabrication of microcapillaries and waveguides for single molecule detection. *Proc. SPIE* **3258**, 141–147 (1998).
- Dittrich, P. S. & Manz, A. Single-molecule fluorescence detection in microfluidic channels — the Holy Grail in μTAS? *Anal. Bioanal. Chem.* **382**, 1771–1782 (2005).

69. Tegenfeldt, J. *et al.* Micro- and nanofluidics for DNA analysis. *Anal. Bioanal. Chem.* **378**, 1678–1692 (2004).
70. Kricka, L. Revolution on a square centimeter. *Nature Biotechnol.* **16**, 513–514 (1998).
71. Tegenfeldt, J. O. *et al.* Stretching DNA in nanochannels. *Biophys. J.* **86**, 596A (2004).
72. Tegenfeldt, J. O. *et al.* The dynamics of genomic-length DNA molecules in 100-nm channels. *Proc. Natl Acad. Sci. USA* **101**, 10979–10983 (2004).
73. Mannion, J. *et al.* Conformational analysis of single DNA molecules undergoing entropically induced motion in nanochannels. *Biophys. J.* **90**, 4538–4545 (2006).
74. Kung, C.-Y. *et al.* Confinement and manipulation of individual molecules in attoliter volumes. *Anal. Chem.* **70**, 658–661 (1998).
75. Han, J. *et al.* Entropic trapping and escape of long DNA molecules at submicron constrictions. *Phys. Rev. Lett.* **83**, 1688–1691 (1999).
76. Cabodi, M. *et al.* Entropic recoil separation of long DNA molecules. *Anal. Chem.* **74**, 5169–5174 (2002).
77. Riehn, R. *et al.* Restriction mapping in nanofluidic devices. *Proc. Natl Acad. Sci. USA* **102**, 10012–10016 (2005).
78. Han, J. & Craighead, H. G. Separation of long DNA molecules in a microfabricated entropic trap array. *Science* **288**, 1026–1029 (2000).
79. Perkins, T. T., Smith, D. E., Larson, R. G. & Chu, S. Stretching of a single tethered polymer in a uniform flow. *Science* **268**, 83–87 (1994).
80. Bakajin, O. B. *et al.* Electrodynamic stretching of DNA in confined environments. *Phys. Rev. Lett.* **80**, 2737–2740 (1998).
81. Service, R. The race for \$1000 genome. *Science* **311**, 1544–1546 (2006).
82. Bialek, W. & Setayeshgar, S. Physical limits to biochemical signaling. *Proc. Natl Acad. Sci. USA* **102**, 10040–10045 (2005).
83. Krapf, D. *et al.* Fabrication and characterization of nanopore-based electrodes with radii down to 2 nm. *Nano Lett.* **6**, 105–109 (2006).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence should be addressed to the author (hgc1@cornell.edu).

Control and detection of chemical reactions in microfluidic systems

Andrew J. deMello¹

Recent years have seen considerable progress in the development of microfabricated systems for use in the chemical and biological sciences. Much development has been driven by a need to perform rapid measurements on small sample volumes. However, at a more primary level, interest in miniaturized analytical systems has been stimulated by the fact that physical processes can be more easily controlled and harnessed when instrumental dimensions are reduced to the micrometre scale. Such systems define new operational paradigms and provide predictions about how molecular synthesis might be revolutionized in the fields of high-throughput synthesis and chemical production.

Ever since Wöhler's laboratory synthesis of urea in 1828 (ref. 1), the chemist's toolkit has predominantly consisted of macroscopic components fabricated from glass. Examples include round-bottomed flasks, test tubes, distillation columns, reflux condensers and retorts. Despite advances in experimental and mechanistic organic chemistry during the past century, it is noteworthy that the basic experimental techniques and associated equipment have remained largely unchanged. There are a number of reasons why traditional synthetic chemistry is performed in the aforementioned equipment, but is there any advantage to performing synthetic chemistry in volumes 5–9 orders of magnitude smaller than those associated with bench-top chemistry? As shown below, the application of techniques cultivated in semiconductor industries have allowed the creation of a new instrumental platform able to efficiently manipulate, process and analyse molecular reactions on the micrometre to nanometre scale. Even at this early stage in the development of 'microfluidic' reaction systems, it is clear that advantages engendered by miniaturization may affect molecular synthesis similarly to the way that the integrated circuit has defined the computer revolution over the past 50 years.

Flow and mixing on the microscale

A primary reason why microfluidic systems provide unusual environments in which to perform synthesis is the dependency of fluid-flow characteristics on scale. Although many diverse effects manifest themselves upon moving from macroscale to microscale environments, some critical features are worthy of discussion.

A tangible effect of reactor miniaturization is that fluid properties become increasingly controlled by viscous forces rather than inertial forces (see page 374). For microfluidic systems, such as blood capillaries, Reynolds numbers (Re) are typically $<10^2$. This represents a situation in which flow is considered essentially laminar, and contrasts with macroscale conduits ($Re > 10^3$) in which flow regimes are almost always turbulent. This behaviour has a direct consequence on mixing within microfluidic systems. Before a reaction between two reagents can occur, intimate contact between the component molecules must be realized through mixing. In its simplest manifestation, this occurs by uniting pure fluid-component streams. Because mixing can only be accomplished by diffusion, rather than through the fast convective processes that dominate in turbulent systems, the only route to

mixing is diffusion across fluidic interfaces (Fig. 1a). Diffusive mixing efficiencies for continuous-flow systems can be measured using the Fourier number, and indicate that mixing timescales increase with the characteristic dimensions of the reactor. Consequently, although mixing via diffusion is inefficient for reactors with characteristic dimensions greater than 1 mm, when diffusion distances drop below 100 μm mixing times can, in theory, become very small.

The ability to controllably and rapidly create a homogenous reactant mixture at the commencement of a reaction is desirable. Indeed, the effect of mixing on the extent of a reaction and product distribution is crucial in reactor design. It is generally recognized that first-order irreversible reactions are not affected by local turbulent mixing, but by the residence time of the system, and conversions can therefore be easily calculated². However, in the case of fast reactions in which two or more reagents are initially present in separate streams, reaction rarely occurs uniformly throughout the whole volume. The rate of reaction is no longer defined by inherent kinetics, but is limited by diffusional rates. Thus, for fast reactions yielding a single product, yield is regarded as a direct measure of the degree of mixing.

The relationship between the reaction rate and the rate of mixing can be reduced to one of three general categories; the chemical regime, the diffusional regime and the mixed chemical/diffusional regime. In the chemical regime, mixing is fast compared with the reaction rate (and is complete before a significant amount of product is generated). In the diffusional regime, reaction is fast, with the rate being limited by the mixing speed. In this case, the reaction rate is independent of the rate constant, and the formation of secondary products in this situation is greatest. Finally, in a mixed chemical/diffusional regime the greatest interaction between chemical reactions and fluid dynamics occurs, and the product distribution depends on both chemical factors (such as reaction kinetics) and diffusional factors (such as the mixing efficiency). To address these variations, a diverse range of microfluidic systems have been designed for the rapid mixing of fluids^{3,4}. They can all be broadly classified as being either passive or active. Passive mixers rely on geometric properties of the channel or fluidic streams to maximize the area over which diffusion can occur, whereas active mixers rely on time-dependent perturbations of the fluid flow to achieve mixing. However, the fact that spatial organization of fluid streams allows mixing to be performed in an extremely rapid and controllable fashion is

¹Electronic Materials Group, Department of Chemistry, Imperial College London, Exhibition Road, South Kensington, London, SW7 2AZ, UK.

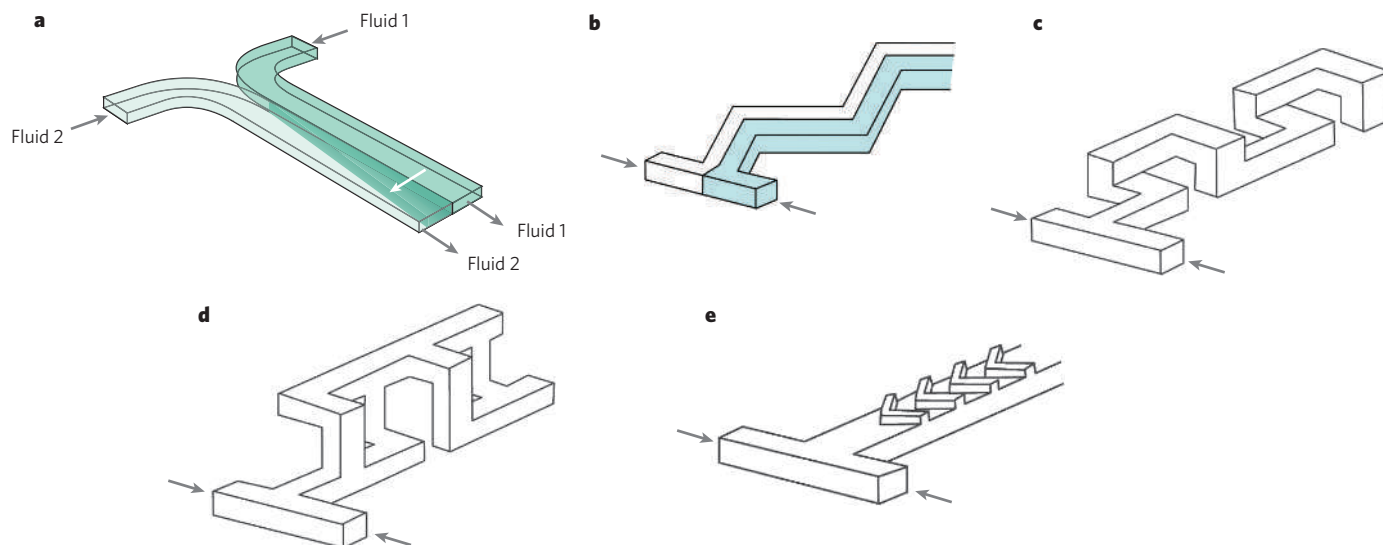


Figure 1 | Microfluidic approaches for mixing in continuous flow. **a**, Mixing of two miscible fluid streams under laminar flow conditions. The component streams mix only by diffusion, creating a dynamic diffusive interface with predictable geometry. **b**, Zigzag-shaped channel for chaotic mixing at high Reynolds numbers⁹. **c**, Three-dimensional L-shaped channel for chaotic mixing at intermediate Reynold numbers¹¹. **d**, Three-dimensional, connected out-of-plane channel for chaotic mixing at intermediate Reynold numbers¹². **e**, Staggered-herringbone grooves for chaotic mixing at low Reynolds numbers¹².

common to all approaches. Both features have significant advantages over macroscale systems and offer potential solutions to a number of key problems faced in contemporary synthesis. These include the ability to probe ultra-fast chemical reactions (with minimal sample consumption), which is beyond the reach of current technologies⁵. An excellent example of such facility was described by Knight *et al.*, who reported a continuous-flow mixer incorporating a hydrodynamic focusing geometry with mixing times of less than 10 μ s and sample consumption rates of nanolitres per second⁶.

Passive mixers have found the widest use in synthetic applications due to their simplicity and operational flexibility. Although operation within laminar-flow regimes can provide rapid mixing if diffusional distances are kept small, in many situations practical limitations (such as minimum feature dimensions) mean that basic flow lamination is inefficient at generating high degrees of mixing within short times. However, rapid mixing with low reagent consumption is achievable using chaotic advection. Put simply, chaotic advection enhances mixing in laminar-flow systems, because it acts to continuously 'stretch' and 'refold' concentrated solute volumes, thereby creating an exponential decrease in striation thickness^{7,8}. Chaotic advection in microfluidic systems can be achieved by introducing obstacles within channels or on channel surfaces, or by modifying channel geometries. In each case, the modification acts to enhance stretching, folding and breaking of the flow. For example, zigzag channels (Fig. 1b) can generate chaotic advection at high Reynolds numbers by recirculation around turns⁹, whereas three-dimensional serpentine channels (Fig. 1c, d) consisting of repeating segments in orthogonal planes can generate chaotic flows at low to intermediate Reynolds numbers^{10,11}. Creation of chaotic flow at low Reynolds numbers has also been established through the use of grooves on channel surfaces (Fig. 1e). A good example of this approach was reported by Strook *et al.*, who used bas-relief, herringbone grooves on a channel bed to induce chaotic mixing at Reynolds numbers between 1 and 100 (ref. 12). More recent studies have also used surface-charge patterning to create electrokinetic mixing in low Reynolds number regimes¹³.

A significant problem encountered in single-phase microfluidic systems is that of achieving rapid and efficient mixing of fluids while minimizing dispersion. Under most circumstances, channel walls impart shear forces on the contained fluid, so under applied hydrodynamic pressure a parabolic velocity profile is established over the cross-section with fluid velocity zero at channel walls and maximum at the centre. The chief implication of this behaviour is that a reaction

mixture sampled after initiation of mixing is formed from an ensemble of volume elements that have spent varying times on-chip. This yields a residence-time distribution that may cause significant variation in the yield, efficiency and product distribution of a reaction. Localization of reagents within discrete droplets is an effective way of eliminating this phenomenon. Several recent studies have exploited the formation of droplets in microfluidic systems to perform a variety of analytical processes^{14–16}. Of particular note are those that use flow instabilities between two immiscible fluids¹⁷. As can be seen in Fig. 2, droplets can be made to form spontaneously when multiple laminar streams of aqueous reagents are injected into an immiscible carrier fluid¹⁸. The formed droplets define picolitre volumes, and because each droplet is isolated from channel surfaces and other droplets, each one acts as an individual reaction vessel. Variation of the cross-sectional dimensions of microchannels can be used to regulate droplet volumes, and flow-rate variation allows control of reagent concentrations¹⁶. Importantly, the use of twisting channel geometries is effective in generating chaotic mixing within droplets, by folding, stretching and reorienting fluid. Consequently, mixing is rapid and reagent transport occurs with no dispersion. Such features, combined with the ability to combine, split and sort droplets, are likely to transform the application of microfluidic systems, and suggest that they would be of use in high-throughput synthesis (due to high sample throughput) and kinetic measurements (due to low sample requirements and negligible dispersion)¹⁹.

Synthetic unit operations

High surface-to-volume ratios are key in defining fluid-flow characteristics at the microscale. Of equal importance is the effect of these ratios on diffusion-mediated mass and heat transfer in reactive processes. For example, typical microfluidic devices exhibit high thermal-transfer efficiencies by virtue of reduced thermal masses and high surface-to-volume ratios, and therefore allow exothermic and/or high temperature reactions to be performed in an efficient and controllable (isothermal) manner. Microfluidic systems have been created to allow efficient biphasic reactions between elemental fluorine and a range of organic substrates. Because the transformation of a carbon–hydrogen bond to a carbon–fluorine bond using fluorine is highly exothermic ($\Delta H = -430 \text{ kJ mol}^{-1}$), safety issues relating to temperature control are of vital importance, especially on a large scale. Indeed, studies by Chambers *et al.* have reported direct fluorination of a range of substrates, including diketones and ketoesters^{20–22}. Other examples

in which microfluidic environments have been shown to provide for efficient temperature and thus reaction control include continuous-flow reactors for multicomponent reactions²³, Swern oxidations²⁴, diazotizations^{25,26}, nitrations²⁷, Andrussow reactions, Reimer–Tiemann formylations²⁸ and carbonylations²⁹. A wide variety of other reactivities have been demonstrated in microfluidic reactor systems, including catalytic hydrogenations and dehydrogenations³⁰, Suzuki couplings³¹, Grubbs metathesis³² and photochemical reactions^{33,34}. As these and many others have been discussed elsewhere^{35–39}, only a small number of recent and illustrative examples are described herein.

In recent years, developments in genomics and proteomics have generated many potential drug targets, each requiring small-molecule modulators. These demands have prescribed massive investment into synthetic technologies that can produce drug candidates on short timescales. The vast majority have used solid-supported chemistry to generate compound libraries.

Unfortunately, the dependence on solid-support technologies has severely limited our ability to perform high-throughput molecule discovery⁴⁰. Put simply, efficient attachment and detachment to and from the support are crucial for successful library generation, increasing the number, time, cost and complexity of process steps and the amounts of required reagents. Moreover, reaction rates of solid-phase reactions are appreciably slower than the corresponding solution-phase processes. Accordingly, the reasons to pursue solution-phase chemistries for library generation are undeniable. The batch nature of conventional solution-phase methods (which use arrays of microwells) is unsuitable for efficient process optimization and high-throughput processing. Conversely, the control of fluidic reagent streams within microfluidic systems allows reactions to be performed within chemical regimes (in which mixing is rapid and reaction timescales are defined by inherent reaction kinetics), whereas operation in a continuous- or segmented-flow format allows compartmentalization and/or spatial identification of multiple reactions⁴¹.

An early demonstration of the use of microfluidics in small-molecule compound-library generation was reported by Mitchell *et al.*, who used distributive mixing of laminar reagent streams to synthesise α -dialkyl-acetamide libraries^{23,42}. Reactions were performed in a serial (time-encoded) or parallel (mass-encoded) fashion, and real-time product identification and quantitation was achieved through integration of the microfluidic reactor with time-of-flight mass spectrometry. Such a combination afforded unprecedented control over the reaction, and real-time identification of the small-molecule products. Further utility of microfluidic systems in making compound libraries has been shown by Garcia-Egido *et al.*, who prepared a series of 2-aminothiazoles by means of a Hantzsch reaction of ring-substituted 2-bromoacetophenones and 1-substituted-2-thioureas⁴³. Additionally, Fernandez-Suarez *et al.* reported automated sequential solution-phase combinatorial synthesis to perform a 2×2 synthesis using the Knoevenagel condensation of 1,3-diketones and aldehydes⁴⁴. The system was configured so as to allow multiple reagent streams to be introduced sequentially under hydrodynamic flow. Subsequent development of this concept allowed the rapid and automated synthesis and analysis of a 7×3 pyrazole library⁴⁵.

In reality, sequential systems are limited in terms of application to high-throughput synthesis. In addition to the possibility of cross-contamination, detection systems are limited to those that can effectively probe small volumes. Consequently, approaches to parallel, solution-phase synthesis in microfluidic reactors have been investigated. Parallel solution-phase synthesis is typically complex and inefficient, requiring multiple reactors and large reagent volumes. Accordingly, the development of microfluidic technologies to efficiently process small volumes of reagents within monolithic devices would represent a significant advance. A step towards integrated, parallel-reaction systems was presented by Kikutani *et al.*⁴⁶. To prove principle, a 2×2 parallel-reaction scheme was transferred to a chip-based format. The primary concern when creating a monolithic system lies in the complex channel topology required to perform reactions in a parallel fashion. Most notably, for an

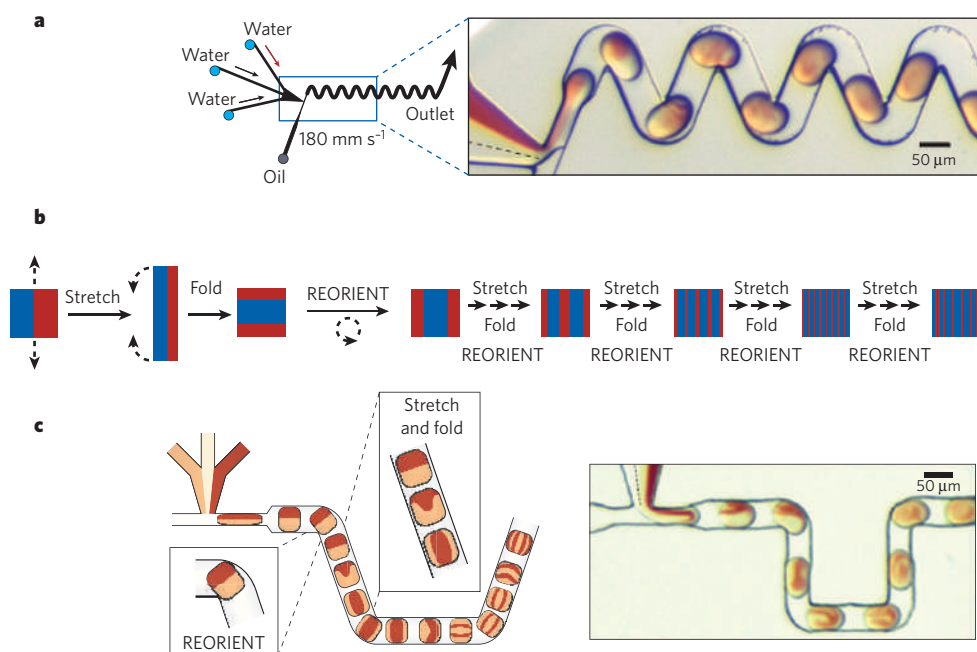


Figure 2 | Formation of microdroplets in microchannels. Microfluidic droplets can be made to spontaneously form when multiple laminar streams of aqueous reagents are injected into an immiscible carrier fluid. The formed droplets define picolitre volumes, and because each droplet is isolated from channel surfaces and other droplets, each one acts as an individual reaction vessel. Chaotic advection within droplets moving through winding channels is used to generate rapid mixing. **a**, Mixing in plugs. Arrow colour corresponds to dye in stream solution. Dashed line in image indicates merging of the streams. (Image reproduced, with permission, from ref. 16.) **b**, Diagram of a fluid element undergoing stretching, folding and reorientation (known as the baker's transformation). Repetition of this process leads to decrease of the striation thickness and facilitates efficient mixing. **c**, Microphotographs of a microfluidic network in which flow patterns inside plugs in the microchannel clearly demonstrate flow patterns reminiscent of the baker's transformation. Red aqueous streams are solutions of $[\text{Fe}(\text{SCN})_x]^{(3-x)+}$ and colourless aqueous streams KNO_3 solution. The oil stream is a solution of water-immiscible fluorinated fluid (perfluorodecalin) with a 10/1 volume/volume ratio of 1H,1H,2H,2H-perfluoro-1-octanol. (Image reproduced, with permission, from ref. 16.)

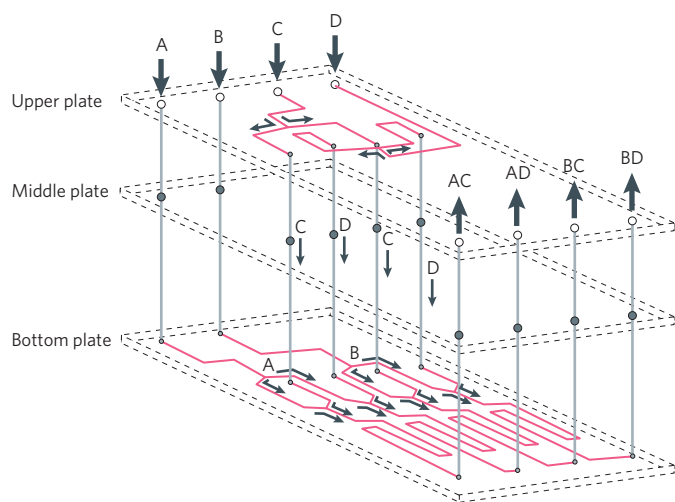


Figure 3 | Schematic view of three-dimensional microchannel circuit for performing parallel combinatorial chemistry. For parallel synthesis, reagents are distributed to each microchannel reactor in a defined manner. There are four inlets and four outlets. Reagent pairs (A and B, and C and D) are libraries of starting molecules. The flow of each reagent is divided into two streams and mixed in four different combinations: AC, AD, BC and BD. An equal distribution of reagents at the branching points is achieved by making the channel length after the branching points equal for both sides⁴⁶.

n by m (compound) combinatorial system in which n and m are greater than 2, a three-dimensional channel network is necessary. To achieve this, two glass substrates were lithographically structured to define two fluidic layers (Fig. 3). Using this approach, a 2×2 combinatorial amide formation reaction was performed with product yields in excess of 90%. Importantly, no impurities or cross-contamination were observed. Although elegant, it is doubtful whether such an approach will ever find application in the synthesis of large libraries due to an associated increase in the complexity of the required fluidic circuitry.

In addition to homogeneous reactions, the large surface-to-volume ratios characteristic of microfluidic reactors provide unique environments for performing heterogeneous chemistry. To this end, Kobayashi *et al.* recently reported enhanced efficiencies of gas-liquid-solid hydrogenation reactions in microchannels⁴⁷. By encapsulating a palladium catalyst in a copolymer matrix attached to the microchannel surface, the metal remains active while irrevocably bound in the solid phase. Using this approach, various substrates (including double bonds, tri-substituted olefins, and triple bonds) were reduced using an annular-flow system. Reactions went to completion within 2 min, and space-time yields were five orders of magnitude higher than equivalent laboratory-scale reactions. The approach is also suitable for large-volume chemical synthesis via 'scale-out', and, importantly, opens up the opportunity for performing a range of catalytic processes at high speed and with negligible catalyst leaching.

Although the above studies provide persuasive arguments for using microfluidic systems in high-throughput synthesis, it is apparent that application is defined by the ability to develop both complex and efficient world-to-chip interfaces, which allow easy coupling between multiple reagent reservoirs and the microfluidic device⁴⁸. At present, reports of microfluidic systems for combinatorial chemistry have, at best, proved principle. Nonetheless, these developments define new paradigms for high-throughput molecular synthesis and provide some of the most credible predictions about how the true power of combinatorial synthesis may be harnessed in molecular discovery.

Enabling nanomaterial synthesis

As I have highlighted, much of the interest in using microfluidic systems for synthetic applications lies in their ability to perform rapid and controllable mixing. This, combined with manipulation of variables

such as temperature, concentration gradients and pressure, dictates that continuous-flow processing on the microscale can be used to synthesize species of specific yet variable characteristics. Perhaps the most interesting demonstration of this feature has been the use of microfluidic reactors to synthesize nanomaterials of defined size and anisotropy. Nanomaterials exhibit optical and electronic properties that depend on their size and shape, and are seen as tailored precursors for functional materials in biological sensing and optoelectronics⁴⁹. These critical dependencies indicate that 'bottom-up' approaches for nanomaterial synthesis must provide for fine control of the physical dimensions of the final product. Synthetic routes involve the particle growth on an atom-by-atom basis, and have been used to create spherical, cubic, tubular and tetrahedral crystallites of well-defined size and shape⁵⁰. Bottom-up approaches have attracted interest owing to their versatility and ease of use, but for many applications deviations about the mean particle diameter must be <1% to achieve the desired selectivity. This is beyond the tolerance of standard macroscale syntheses, and it is almost always necessary to use some form of post-treatment (including chromatography, sedimentation, precipitation and photocorrosion) to extract the desired particle size⁵¹. Accordingly, nanoparticles with narrow size distributions can be extracted, but because the starting point for all such methods is a polydisperse sample, product yields are low.

An ideal recipe for nanoparticle synthesis must ensure that nucleation of solute molecules (to form 'seed' particles) occurs on a timescale that is short compared with the characteristic growth time (in which the seeds capture dissolved solutes). Moreover, nucleation and growth should occur in an environment in which chemical state functions are precisely controlled⁵². If these conditions are not met, the size of critical nuclei and growth rates will vary according to location, and result in a distribution of particle sizes. In many respects, microfluidic systems provide an ideal medium for nanoparticle production. Because both mass and thermal transfer are rapid, temperatures may be defined with precision or varied on short timescales. Additionally, reagents can be rapidly and efficiently mixed to ensure homogeneous reaction environments, while allowing for additional reagents to be added at predefined times.

Recent studies have demonstrated that microfluidic reactors drastically outperform macroscale systems in the direct production of nanoparticles. Using simple flow regimes whereby component streams are mixed at low Reynolds numbers and in continuous flow, variations in reaction residence times, temperatures and reagent concentrations are used to control average particle size, while sample size distributions are minimized through a reduction in residence-time distributions and precise control of chemical state functions. In this way, high-quality cadmium sulphide^{53,54}, cadmium selenide^{55,56}, palladium⁵⁷, silver^{58,59}, gold⁶⁰, copper⁶¹, titania⁶² and CdSe-ZnS core-shell nanoparticles⁶³ have all been synthesized directly.

Recent studies have addressed the issue of further minimizing particle size distributions through the development of segmented-flow reactors. Shestopalov *et al.* demonstrated a two-step chemical synthesis of colloidal CdS and CdS-CdSe core-shell nanoparticles in a droplet-based microreactor⁶⁴. Importantly, the system affords millisecond time control and also allows the stages of a multistep reaction to be initiated at precise times. In addition, Chan *et al.* have reported the use of microfluidic-droplet reactors for the high-temperature synthesis of CdSe nanoparticles⁶⁵, whereas Yen *et al.* have used gas-liquid segmented-flow reactors containing multiple temperature zones for the synthesis of high quality CdSe quantum dots⁶⁶ (Fig. 4). In all of these studies, enhanced mixing and reduced residence-time distributions fuelled the improvements in yield and size distribution.

Finally, it should be noted that microfluidic systems can be used to create higher-order nanostructures that are inaccessible via conventional methods. Millman *et al.* recently reported the synthesis of anisotropic particles in static-microdroplet reactors⁶⁷. Droplets with volumes between 500 and 2,000 nl were floated on the surface of a perfluorinated oil. Because droplets can be trapped and manipulated by electrical fields generated by electrode arrays, droplets containing suspensions of

nanoparticles and polymers can be induced to form complex particle structures. For example, 'striped' multilayer particles could be generated from ternary mixtures of gold, fluorescent latex and silica particles, and core-shell particles could be synthesized by encapsulation of dried supraparticles or droplets of aqueous suspension inside polymer shells.

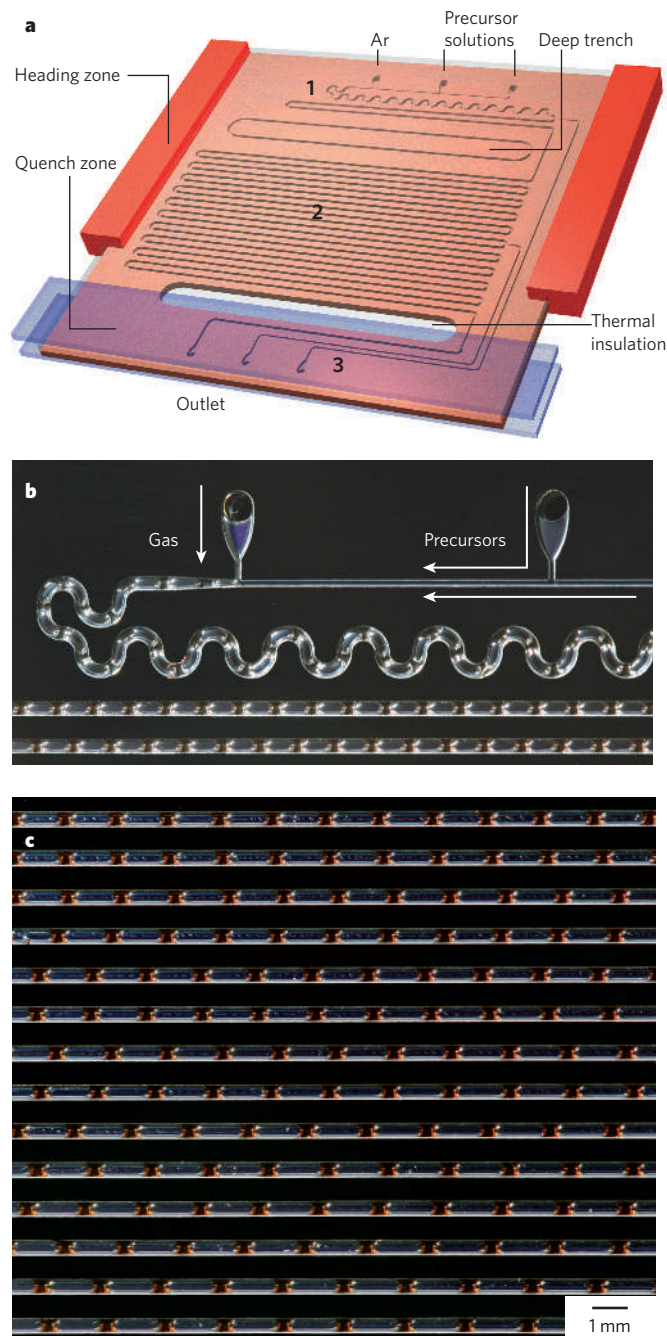


Figure 4 | Microfluidic reactor for nanoparticle production. **a**, The reactor allows rapid precursor mixing (sector 1), controlled particle growth (sector 2) and reaction quenching (sector 3). The reactor accommodates a ~1-metre-long reaction channel and two shallow side channels for collecting reaction aliquots. A halo etch region allows localization of temperature zones for reaction ($>260^{\circ}\text{C}$) and quenching ($<70^{\circ}\text{C}$) on the device. Precursor solutions are delivered into the heated section separately, and an argon (Ar) gas stream generates segmented gas-liquid flow. Recirculation within the liquid slugs (droplets) rapidly mixes reagents and initiates the reaction. The reaction is stopped when the fluids enter the cooled outlet region of the device. **b**, **c**, photographs of heated inlets (**b**) and main channel section (**c**). Red segments show the reaction solution; dark segments define Ar gas; $T = 260^{\circ}\text{C}$; gas flow rate = $60\ \mu\text{l min}^{-1}$; liquid flow rate = $30\ \mu\text{l min}^{-1}$. (Images reproduced, with permission, from ref. 66.)

Microfluidic reactors for DNA amplification

The use of microfluidic systems in synthesis is not confined to small-molecule or nanoparticle synthesis. Indeed, many early applications of microfluidics focused on biological reactions^{68–71}. The most investigated biological reaction in microfluidic systems is DNA amplification via the polymerase chain reaction (PCR). This enzyme-catalysed reaction allows any nucleic-acid sequence to be generated in abundance *in vitro*, and has become a fundamental tool in molecular biology⁷². Although simple to implement, PCR in conventional thermal cyclers is slow and inefficient due to large thermal masses associated with instrumentation. To address this issue, many microfabricated devices for PCR have been reported, with performance gains achieved through a reduction in the thermal mass of the system. Effective and rapid PCR in volumes ranging from 1 μl to 50 μl has been performed using various heating mechanisms, including infrared-mediated thermal cycling⁷³, microwave heating⁷⁴, Joule heating⁷⁵ and resistive heating⁷⁶.

Importantly, the use of micromachining methods has also engendered new approaches to PCR. For example, Krishnan *et al.* have presented an elegant microfluidic system for PCR that relies on the control of thermal convection in a Rayleigh–Bénard cell to provide thermal cycling conditions⁷⁷. Moreover, the widespread adoption of continuous-flow modalities for PCR has been facilitated through the use of microfabricated systems. Continuous-flow PCR (in which a sample is moved continuously through multiple reaction zones held at specific temperatures) has been shown to provide for ultra-fast DNA amplification. Originally described by Kopp *et al.*⁷⁸, this approach has yielded the fastest reaction times to date, as the small-volume fluid elements can be heated or cooled to the required temperature within a few milliseconds. More recently, this concept has been extended to create integrated systems for performing reverse transcription and PCR within a single microdevice^{79,80}.

Finally, it should be noted that, recently, much interest has focused on the creation of highly integrated microfluidic systems for complex biological processing. An elegant example of an integrated monolithic device for DNA amplification was reported by Liu *et al.*⁸¹. An integrated microfluidic device consisting of mixer elements, fluidic valves and pumps, microchannels, chambers, heaters, and microarray sensors allows for sequential sample preparation (such as cell pre-concentration, purification and cell lysis), PCR, DNA hybridization and electrochemical detection. Moreover, Lagally *et al.* have described the refinement of an integrated genetic-analysis microsystem for PCR and capillary electrophoresis⁸². The microdevice contains microfabricated heaters, temperature sensors and membrane valves to provide controlled sample manipulation and processing of DNA within 200-nl PCR chambers. Using the system, DNA amplification and product sizing could be performed within 10 min, and its utility established through pathogen detection, and genotyping directly from whole *Escherichia coli* and *Staphylococcus aureus* cells.

In general, it is fair to say that microfluidic approaches for PCR have delivered, in terms of the expected advantages with respect to macroscale systems. Indeed, a number of the current generation of commercial thermal cyclers for PCR have embraced the basic tenets of miniaturization. However, the use of microfabricated systems in real-world applications (such as medical diagnostics) is yet to become reality and will ultimately be defined by the ability to create highly integrated microfluidic systems that can handle and process complex biological fluids, and be manufactured at low cost. Fortunately, progress is being made in this crucial area, with contemporary examples of highly-integrated microsystems demonstrating that complex biological processing (such as Sanger DNA sequencing) can be performed at higher speeds and with superior efficiencies than previously achievable⁸³ (Fig. 5).

Extracting information at the microscale

Compared with the macroscale, microfluidic systems engender significant advantages in terms of speed, throughput, yield, selectivity and control. All are directly facilitated by system downscaling and associated improvements in mass and thermal transfer. Nevertheless, manipulation and processing of samples with instantaneous volumes

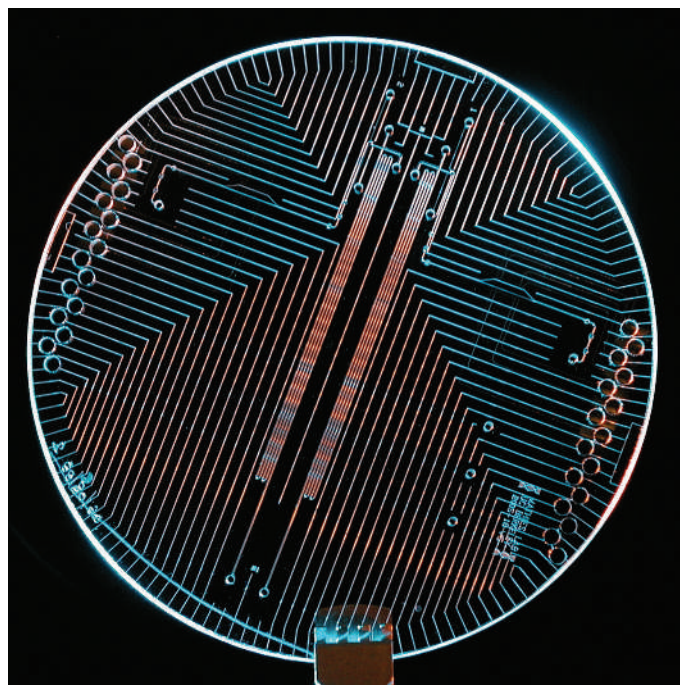


Figure 5 | Integrated microfluidic bioprocessor. A nanolitre-scale microfabricated bioprocessor that integrates thermal cycling, sample purification and capillary electrophoresis for Sanger sequencing. The hybrid glass–poly(dimethylsiloxane) microdevice contains 250-nl reactors, affinity-capture purification chambers, high-performance capillary electrophoresis channels, and pneumatic valves and pumps. Such integration enables complete Sanger sequencing from only 1 fmol of DNA template. (Image reproduced, with permission, from ref. 83.)

ranging from a few picolitres to hundreds of nanolitres provides a significant challenge for analyte detection and identification, and in many ways defines the principal limitations of current microfluidic systems. Detailed evaluation of detection methods for small-volume environments are provided elsewhere^{84,85}, however, effective detection within microfluidic environments is clearly defined by a close interrelationship of factors such as detector sensitivity, response times, detection limits and information content. Crucially, although microfluidic systems have been shown to be highly effective at generating conditions in which variables such as reagent concentration^{86,87}, temperature⁸⁸ and pH⁸⁹ can be controlled with precision, extraction of the available information is normally non-ideal. In other words, although microfluidic reactors generate high-quality chemical information, detection protocols are often inefficient in extracting all available information. For example, we have seen that segmented flow within microfluidic channels allows generation of picolitre-sized droplets (of variable chemical composition) at frequencies in excess of 100 Hz. This means that thousands of individual reactions can be processed in very short times. However, few — if any — studies have successfully exploited this feature, with analytical throughput being defined by the speed at which the detection system can operate.

Despite the challenges of information extraction, some recent reports demonstrate developments in high-throughput chemistry and screening. Ratner *et al.* have reported the use of a microfluidic reaction system to systematically study glycosylation reactions through control of reaction times and temperatures⁹⁰. Using such an approach, optimal temperature/concentration/flow rate protocols were established by in-line high-performance liquid chromatography (HPLC) analysis of the chip effluent. Interestingly, the mannosylation of 2,3,4-tri-O-benzyl-methyl mannoside was achieved in optimal yield at a temperature of -60°C and a reaction time of 213 s. However, via rapid reaction screening, an almost comparable performance was achieved at a temperature of -35°C and a reaction time of 26 s. In a similar manner, Leung *et al.* reported

ultra-fast screening of reaction conditions within a continuous-flow micromixer using confocal Raman microscopy⁹¹. The catalytic oxidation of isopropyl alcohol to acetone using tetra-*n*-propylammonium perruthenate and *N*-methylmorpholine-*N*-oxide was assessed as a function of reagent concentrations and residence times. The composition of the reaction effluent was easily determined, and information relating to catalyst/co-oxidant ratios, catalyst turnovers and reaction endpoints extracted. In both studies, the extraction of chemical information in short times, using minimal reagent volumes provides a direct route to rapid synthetic-process optimization, which is not possible in macroscale environments. Very recently, Hatakeyama *et al.* also illustrated rapid optimization of organic reactions within a capillary-based microdroplet reactor⁹². Deacetylation reactions were performed in high-throughput by creating droplet reactors surrounded by a fluorinated carrier fluid, and merging these with a substrate flow. Resulting plugs form, flow into receiving tubing, are stopped, and are isolated by sealing. After incubation, sample is then deposited onto a matrix-assisted laser desorption/ionization (MALDI) plate and analysed by MALDI-mass spectrometry (MALDI-MS) to assess reaction progress. Reaction-condition screening was achieved by screening more than 40 reagents to evaluate reactivity, and then repeating the screen with a more focused reagent set while varying reaction conditions (such as reaction time, solvent and concentration). Although the optimization timescale is limited by the speed of MALDI-MS, the system is simple to implement and able to process reactions on a scale of less than 1 μg per reaction.

Microfluidic factories

Most applications of microfluidic systems in synthesis have focused on the implementation of individual reaction units to demonstrate enhanced performance characteristics compared with macroscale systems. However successful (in terms of yield, purity or speed) these systems are at generating a product, their use in production applications is determined by the ease with which they can be used to generate significant volumes of product in short times. Although characterized by instantaneous volumes in the nanolitre range, microfluidic systems can be configured to achieve such a goal. For example, a system generating product at a concentration of 10% at a flow rate of $200\ \mu\text{l min}^{-1}$ will yield 1.2 ml of product in 1 hour. Therefore, 100 reactors operating in parallel will produce $120\ \text{ml h}^{-1}$, a rate comparable to that of many fine chemical processes. This simple calculation is based on typical examples of reactions with low to moderate yield, and demonstrates that fine-scale processes can be simulated on chip arrays that are within the bounds of current technological development.

A recent example of this concept was provided by Chambers *et al.*, who reported effective scale-out of a steel microfluidic reactor for direct fluorination⁹³. Using an array of parallel microchannels coupled by a simple manifold interface to reagent flows, high-purity fluorinated products could be generated at rates of $30\ \text{g day}^{-1}$. Concurrent operation of ten such reactors mimics a small pilot plant operation, running under the same conditions as the laboratory synthesis. Importantly, the system is easy to maintain, operates continuously, requires single-source fluid delivery and operates under safe conditions. In addition, Lee *et al.* recently reported a highly integrated microfluidic system for the direct synthesis of molecular imaging probes used in positron emission tomography (PET)⁹⁴. Although the required volumes are relatively small, the challenge associated with producing radiolabelled chemicals relates to production of highly pure materials within very short timescales (due to the half-life of [^{18}F]fluorine being 110 min). Using a highly integrated microfluidic (Fig. 6), [^{18}F]fluoride concentration, water evaporation, radiofluorination, solvent exchange and hydrolytic deprotection were performed rapidly and with high radio-chemical yield to synthesize 2-deoxy-2- ^{18}F fluoro-D-glucose. Significantly, product was generated in high enough yield to be used in *in vivo* PET studies.

An elegant approach to 'scale-out' has also been described by Kikutani *et al.*⁹⁵. Simulation of large-scale flows was achieved using a 'pile-up' reactor in which ten glass microchannel circuits were integrated (via thermal bonding) to form a single glass structure. Importantly, although

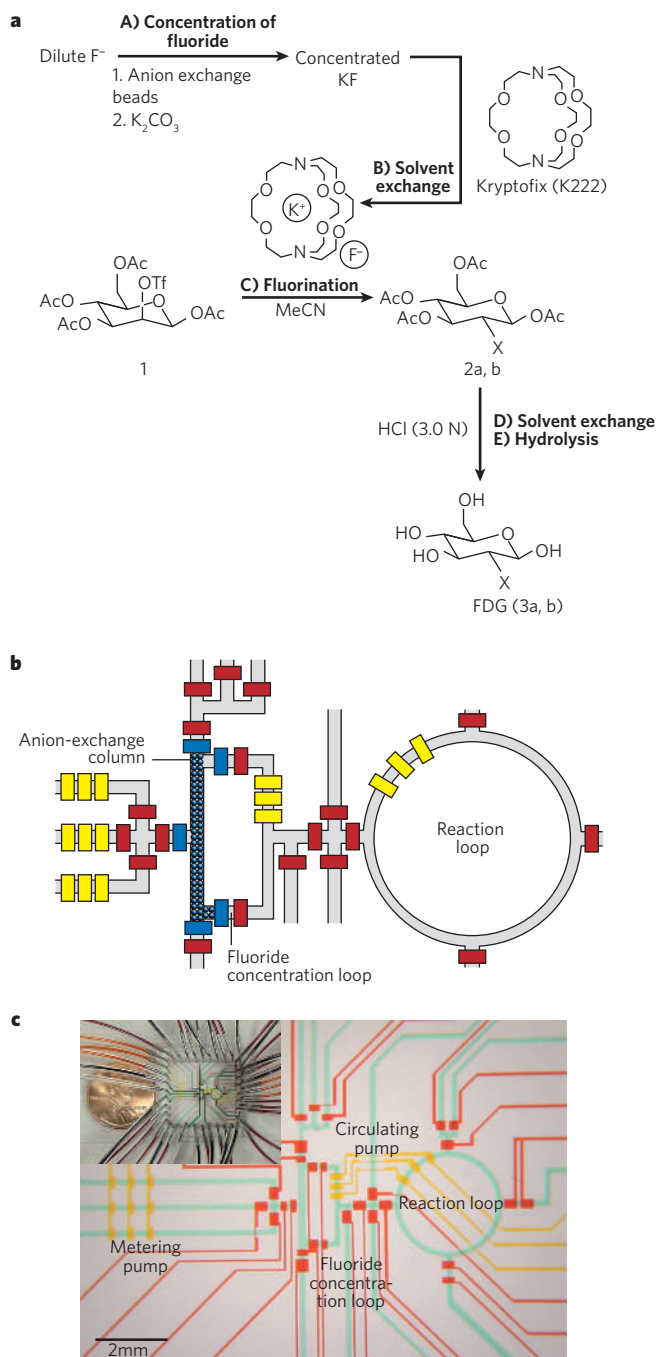


Figure 6 | Schematic representation of a chemical reaction circuit used to synthesize 2-deoxy-2-fluoro-D-glucose. **a**, Five sequential processes are shown. **A**, Concentration of dilute fluoride ion with an anion-exchange column located in a rectangle-shaped fluoride concentration loop. **B**, Solvent exchange from water to dry MeCN. **C**, Fluorination of the D-mannose triflate precursor 1. X represents ^{18}F (2a and 3a) or ^{19}F (2b and 3b). **D**, Solvent exchange back to water. **E**, Acidic hydrolysis of the fluorinated intermediate 2a (or 2b) in a ring-shaped reaction loop. Nanogram amounts of 2-deoxy-2-fluoro-D-glucose (FDG) (3a, b) define the final product. **b**, The operation of the circuit is controlled by pressure-driven valves, with their delegated responsibilities illustrated by their colours: red, regular valves (for isolation); yellow, pump valves (for fluidic metering circulation); blue, sieve valves (for trapping anion exchange beads in the column module). (Image adapted, with permission, from ref. 94.) **c**, Optical micrograph of the central area of the circuit. The various channels have been loaded with food dyes to help visualize the different components of the microfluidic chip; colours are as in **a**, plus green for fluidic channels. Inset, actual view of the device; a US one-cent coin (diameter 18.9 mm) is shown for comparison. (Image reproduced, with permission, from ref. 94.)

the maximum throughput for the ten-layered pile-up reactor was ten times larger than that of a single microfluidic device, the reaction yield was maintained at 80%. Moreover, a reaction productivity of a few grams per hour scale could be maintained, comparing well with many fine-chemical processes.

After almost a decade of intensive study on the utility of microfluidic systems in chemical production, a few observations can be made. Early research has been successful in demonstrating that many fundamental synthetic transformations can be performed with improved space–time yields, selectivities, reaction residence times and conversions with microfluidics compared with traditional methods. However, application of such systems in industrial environments requires a better understanding of other parameters, such as scalability, facile process control, safety, profitability and operational flexibility⁹⁶. Although there are some examples in commercial production, industry has been slow to embrace microfluidic innovations. This is due in part to limited long-term data on the performance and control of microfluidic reactors in operational environments, and also to cultural factors, such as the widespread investment in and deployment of batch reactors. Indeed, because microfluidic reaction systems can be used to good effect in a number of the stages involved in a chemical process (for example, in compound screening, laboratory-scale process development and production), the extensive implementation of microfluidic factories for chemical production may ultimately be defined by their acceptance as *de facto* tools in these upstream processes, which will, in turn, make them a more natural choice as an industrial tool.

Outlook

The success of microfluidic systems in molecular synthesis is due largely to the exploitation of atypical fluid behaviour in small-volume environments. The fact that fluid properties become increasingly controlled by viscous forces as reaction volumes are reduced dictates that mixing can only be accomplished through diffusion. Nevertheless, at this scale diffusion provides a driver for both rapid and controlled mixing of fluids. Various continuous-flow and batch microfluidic reactors have used these basic ideas to good effect and demonstrated performance characteristics superior to macroscale systems. Although such gains are indisputable, it is less clear how microfluidics will ultimately affect synthesis in both research and industrial environments. Most studies that have used microfluidics in synthesis have either demonstrated the transferral of unit operations from conventional to chip-based formats, or have verified the reality of performance enhancements predicted by theory. This state of affairs defines the initial phase in the lifecycle of microfluidic technology. In many ways, the secondary and critical phase has already begun, whereby microfluidic tools are developed and refined to address and solve fundamental questions. One of the most visible examples of this progression has been the development of segmented-flow droplet reactors. As we have seen, the use of such systems as a basic tool in addressing high-throughput screening and kinetic studies of complex chemical and biological systems defines a totally new approach.

Moreover, as discussed, perhaps the biggest challenge is the efficient extraction and utilization of the vast amounts of information produced. This dictates the development and integration of sensitive detection systems that can process significant amounts of information per unit time. For example, it is expected that the use of high-performance microfluidics will undoubtedly impact the field of catalysis. Droplet systems could be configured to contain a gene and a transcription/translation system that includes all the required ingredients for *in vitro* protein expression. The selection of catalytically-active enzymes can then be performed using molecular biology to generate new catalysts and analytical techniques with which to screen them. The power of such a discovery platform is such that it should be able to screen catalysts at rates up to five orders of magnitude faster than is possible at present.

In conclusion, molecular synthesis in both chemistry and biology has focused, and will continue to focus, on high-throughput experimentation on small samples. Progress in disciplines such as genomics,

proteomics, drug-discovery and high-throughput screening requires new and robust tools that will enable the extraction of enormous amounts of information and, in turn, provide the basis of a better understanding of chemical and biological phenomena. Microfluidics may just provide such tools. ■

- Wöhler, F. *Grundriß der Organischen Chemie* (Duncker und Humblot, Berlin, 1848).
- Levenspiel, O. *Chemical Reaction Engineering* (Wiley & Sons, Chichester, 1999).
- Nguyen, N. T. & Wu, Z. Micromixers — a review. *J. Micromech. Microeng.* **15**, R1–R16 (2005).
- Ottino, J. M., Wiggins, S. Introduction: mixing in microfluidics. *Phil. Trans. R. Soc. Lond. A* **362**, 923–935 (2004).
- Shastri, M. C. R., Luck, S. D. & Roder, H. A Continuous-flow capillary mixing method to monitor reactions on the microsecond time scale. *Biophys. J.* **74**, 2714–2721 (1998).
- Knight, J. B., Vishwanath, A., Brody, J. P. & Austin, R. H. Hydrodynamic focusing on a silicon chip: mixing nanoliters in microseconds. *Phys. Rev. Lett.* **80**, 3863–3866 (1998).
- Ottino, J. M. The mixing of fluids. *Sci. Am.* **260**, 56–67 (1989).
- Ottino, J. M. Mixing, chaotic advection, and turbulence. *Ann. Rev. Fluid Mech.* **22**, 207–254 (1990).
- Mengeaud, V., Josserand, J. & Girault, H. H. Mixing processes in a zigzag microchannel: finite element simulation and optical study. *Anal. Chem.* **74**, 4279–4286 (2002).
- Chen, H. & Meiners, J. C. Topologic mixing on a microfluidic chip. *Appl. Phys. Lett.* **84**, 2193–2195 (2004).
- Liu, R. H. *et al.* Passive mixing in a three-dimensional serpentine microchannel. *J. Microelectromech. Syst.* **9**, 190–197 (2000).
- Stroock, A. D. *et al.* Chaotic mixer for microchannels. *Science* **295**, 647–651 (2002).
- Biddiss, E., Erickson, D. & Li, D. Q. Heterogeneous surface charge enhanced micromixing for electrokinetic flows. *Anal. Chem.* **76**, 3208–3213 (2004).
- Burns, M. A. *et al.* An integrated nanoliter DNA analysis device. *Science* **282**, 484–487 (1998).
- Taniguchi, T., Torii, T. & Higuchi, T. Chemical reactions in microdroplets by electrostatic manipulation of droplets in liquid media. *Lab Chip* **2**, 19–23 (2002).
- Bringer, M. R., Gerds, C. J., Song, H., Tice, J. D. & Ismagilov, R. F. Microfluidic systems for chemical kinetics that rely on chaotic mixing in droplets. *Phil. Trans. R. Soc. Lond. A* **362**, 1087–1104 (2004).
- Tice, J. D., Song, H., Lyon, A. D. & Ismagilov, R. F. Formation of droplets and mixing in multiphase microfluidics at low values of the Reynolds and the capillary numbers. *Langmuir* **19**, 9127–9133 (2003).
- Song, H., Tice, J. D. & Ismagilov, R. F. A microfluidic system for controlling reaction networks in time. *Angew. Chem. Int. Edn* **42**, 768–772 (2003).
- Song, H. & Ismagilov, R. F. Millisecond kinetics on a microfluidic chip using nanoliters of reagents. *J. Am. Chem. Soc.* **125**, 14613–14619 (2003).
- Chambers, R. D. & Spink, R. C. H. Microreactors for elemental fluorine. *Chem. Commun.* 883–884 (1999).
- Chambers, R. D., Holling, D., Spink, R. C. H. & Sandford, G. Elemental fluorine Part 13. Gas-liquid thin film microreactors for selective direct fluorination. *Lab Chip* **1**, 132–137 (2001).
- Chambers, R. D., Fox, M. A. & Sandford, G. Elemental fluorine Part 18. Selective direct fluorination of 1,3-ketoesters and 1,3-diketones using gas/liquid microreactor technology. *Lab Chip* **5**, 1132–1139 (2005).
- Mitchell, M. C., Spikmans, V. & deMello, A. J. Microchip-based synthesis and analysis: control of multicompartment reaction products and intermediates. *Analyst* **126**, 24–27 (2001).
- Kawaguchi, T., Miyata, H., Ataka, K., Mae, K. & Yoshida, J. Room-temperature Swern oxidations by using a microscale flow system. *Angew. Chem. Int. Edn* **44**, 2413–2416 (2005).
- Salimi-Moosavi, H., Tang, T. & Harrison, D. J. Electroosmotic pumping of organic solvents and reagents in microfabricated reactor chips. *J. Am. Chem. Soc.* **119**, 8716–8717 (1997).
- Wootton, R. C. R., Fortt, R. & de Mello, A. J. On chip generation and reaction of unstable intermediates — monolithic nanoreactors for diazonium chemistry: azo dyes. *Lab Chip* **2**, 5–7 (2002).
- Ducry, L. & Roberge, D. M. Controlled autocatalytic nitration of phenol in a microreactor. *Angew. Chem.* **117**, 8186–8189 (2005).
- Iles, A., Fortt, R. & de Mello, A. J. Thermal optimisation of the Reimer-Tiemann reaction using thermochromic liquid crystals on a microfluidic reactor. *Lab Chip* **5**, 550–554 (2005).
- Miller, P. W. *et al.* Rapid formation of amides via carbonylative coupling reactions using a microfluidic device. *Chem. Commun.* 546–548 (2006).
- Besser, R. S., Ouyang, X. & Surangelikar, H. Hydrocarbon hydrogenation and dehydrogenation reactions in microfabricated catalytic reactors. *Chem. Eng. Sci.* **58**, 19–26 (2003).
- Greenway, G. M., Haswell, S. J., Morgan, D. O., Skelton, V. & Styring, P. The use of a novel microreactor for high throughput continuous-flow organic synthesis. *Sens. Actuators B* **63**, 153–158 (2000).
- Schwesinger, N., Pieper, G. & Wurziger, H. Method for carrying out a metathesis reaction of unsaturated organic compounds. World Patent WO0170387 (2001).
- Lu, H., Schmidt, M. A. & Jensen, K. F. Photochemical reactions and on-line UV detection in microfabricated reactors. *Lab Chip* **1**, 22–28 (2001).
- Wootton, R. C. R., Fortt, R. & de Mello, A. J. A microfabricated nanoreactor for safe, continuous generation and use of singlet oxygen. *Org. Proc. Res. Dev.* **6**, 187–189 (2002).
- Watts, P. & Haswell, S. J. Continuous-flow reactors for drug discovery. *Drug Discov. Today* **8**, 586–593 (2003).
- Jähnisch, K., Hessel, V., Löwe, H. & Baerns, M. Chemistry in microstructured reactors. *Angew. Chem. Int. Edn* **43**, 406–446 (2004).
- Brivio, M., Verboom, W. & Reinhoudt, D. N. Miniaturized continuous-flow reaction vessels: influence on chemical reactions. *Lab Chip* **6**, 329–344 (2006).
- Watts, P. & Haswell, S. J. The application of microreactors for small-scale organic synthesis. *Chem. Eng. Technol.* **28**, 290–301 (2005).
- Jensen, K. F. Microreaction engineering — is small better? *Chem. Eng. Sci.* **56**, 293–303 (2001).
- Geysen, H. M., Schoonen, F., Wagner, D. & Wagner, R. Combinatorial compound libraries for drug discovery: an ongoing challenge. *Nature Rev. Drug Discov.* **2**, 222–230 (2003).
- Dittrich, P. S. & Manz, A. Lab-on-a-chip: microfluidics in drug discovery. *Nature Rev. Drug Discov.* **5**, 211–218 (2006).
- Mitchell, M. C., Spikmans, V., Manz, A. & deMello, A. J. Microchip-based synthesis and total analysis system (μSYNTAS): chemical microprocessing for generation and analysis of compound libraries. *J. Chem. Soc. Perkin Trans.* 1514–518 (2001).
- Garcia-Egido, E., Wong, S. Y. F. & Warrington, B. H. A Hantzsch synthesis of 2-aminothiazoles performed in a heated microreactor system. *Lab Chip* **2**, 31–33 (2002).
- Fernandez-Suarez, M., Wong, S. Y. F. & Warrington, B. H. Synthesis of a three-member array of cycloadducts in a glass microchip under pressure driven flow. *Lab Chip* **2**, 170–174 (2002).
- Garcia-Egido, E., Spikmans, V., Wong, S. Y. F. & Warrington, B. H. Synthesis and analysis of combinatorial libraries performed in an automated micro reactor system. *Lab Chip* **3**, 73–76 (2003).
- Kikutani, Y. *et al.* Glass microchip with three-dimensional microchannel network for 2 x 2 parallel synthesis. *Lab Chip* **2**, 188–192 (2002).
- Kobayashi, J. *et al.* A microfluidic device for conducting gas-liquid-solid hydrogenation reactions. *Science* **304**, 1305–1308 (2004).
- Fredrickson, C. K. & Fan, Z. H. Macro-to-micro interfaces for microfluidic devices. *Lab Chip* **4**, 526–533 (2004).
- Alivisatos, A. P. Semiconductor clusters, nanocrystals, and quantum dots. *Science* **271**, 933–937 (1996).
- LaMer, V. K. & Dinegar, R. H. Theory, production and mechanism of formation of monodispersed hydrosols. *J. Am. Chem. Soc.* **72**, 4847–4854 (1950).
- Murray, C. B., Kagan, C. R. & Bawendi, M. G. Synthesis and characterization of monodisperse nanocrystals and close-packed nanocrystal assemblies. *Annu. Rev. Mater. Sci.* **30**, 545–610 (2000).
- deMello, J. & deMello, A. Microscale reactors: nanoscale products. *Lab Chip* **4**, 11N–15N (2004).
- Edel, J. B., Fortt, R., deMello, J. C. & deMello, A. J. Microfluidic routes to the controlled production of nanoparticles. *Chem. Commun.* 1136–1137 (2002).
- Hung, L.-H. *et al.* Alternating droplet generation and controlled dynamic droplet fusion in microfluidic device for CdS nanoparticle synthesis. *Lab Chip* **6**, 174–178 (2006).
- Krishnasadan, S., Tovilla, J., Vilar, J., deMello, A. J. & deMello, J. C. On-line analysis of CdSe nanoparticle formation in a continuous-flow chip-based microreactor. *J. Mater. Chem.* **14**, 2655–2660 (2004).
- Yen, B. K. H., Stott, N. E., Jensen, K. F. & Bawendi, M. G. A continuous-flow microcapillary reactor for the preparation of a size series of CdSe nanocrystals. *Adv. Mater.* **15**, 1858–1862 (2003).
- Song, Y., Kumar, C. S. S. R. & Holmes, J. Synthesis of Pd nanoparticles using a continuous-flow polymeric micro reactor. *J. Nanosci. Nanotech.* **4**, 788–793 (2004).
- He, S. *et al.* Effects of interior wall on continuous fabrication of silver nanoparticles in microcapillary reactor. *Chem. Lett.* **34**, 748–749 (2005).
- Lin, X. Z., Terepka, A. D. & Yang, H. Synthesis of silver nanoparticles in a continuous-flow tubular microreactor. *Nano Lett.* **4**, 2227–2232 (2004).
- Wagner, J., Kirner, T., Mayer, G., Albert, J. A. & Köhler, J. M. Generation of metal nanoparticles in a microchannel reactor. *Chem. Eng. J.* **101**, 251–260 (2004).
- Song, Y. *et al.* Investigations into sulfobetaine-stabilized Cu nanoparticle formation: toward development of a microfluidic synthesis. *J. Phys. Chem. B* **109**, 9330–9338 (2005).
- Wang, H., Nakamura, H., Uehara, M., Miyazaki, M. & Maeda, H. Preparation of titania particles utilizing the insoluble phase interface in a microchannel reactor. *Chem. Commun.* 1462–1463 (2002).
- Wang, H. *et al.* Continuous synthesis of CdSe–ZnS composite nanoparticles in a microfluidic reactor. *Chem. Commun.* 48–49 (2004).
- Shestopalov, I., Tice, J. D. & Ismagilov, R. F. Multi-step synthesis of nanoparticles performed on millisecond time scale in a microfluidic droplet-based system. *Lab Chip* **4**, 316–321 (2004).
- Chan, E. M., Alivisatos, A. P. & Mathies, R. A. High-temperature microfluidic synthesis of CdSe nanocrystals in nanoliter droplet. *J. Am. Chem. Soc.* **127**, 13854–13861 (2005).
- Yen, B. K. H., Günther, A., Schmidt, M. A., Jensen, K. F. & Bawendi, M. G. A microfabricated gas-liquid segmented flow reactor for high-temperature synthesis: the case of CdSe quantum dots. *Angew. Chem. Int. Edn* **44**, 5447–5451 (2005).
- Millman, J. R., Bhatt, K. H., Prevost, B. G. & Velev, O. D. Anisotropic particle synthesis in dielectrophoretically controlled microdroplet reactors. *Nature Mater.* **4**, 98–102 (2004).
- Magally, E. T. & Mathies, R. A. Integrated genetic analysis microsystems. *J. Phys. D* **37**, R245–R261 (2004).
- Beebe, D. J., Mensing, G. A. & Walker, G. M. Physics and applications of microfluidics in biology. *Annu. Rev. Biomed. Eng.* **4**, 261–286 (2002).
- Auroux, P.-A., Koc, Y., deMello, A., Manz, A. & Day, P. J. R. Miniaturised nucleic acid analysis. *Lab Chip* **4**, 534–546 (2004).
- Jakeway, S. C., de Mello, A. J. & Russell, E. Miniaturized total analysis systems for biological analysis. *Fresenius J. Anal. Chem.* **366**, 525–539 (2000).
- Mullis, K. B. *et al.* Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* **51**, 263–273 (1986).
- Oda, R. P. *et al.* Infrared-mediated thermocycling for ultrafast polymerase chain reaction amplification of DNA. *Anal. Chem.* **70**, 4361–4368 (1998).
- Fermér, C., Nilsson, P. & Larhed, M. Microwave-assisted high-speed PCR. *Eur. J. Pharm. Sci.* **18**, 129–132 (2003).
- Hu, G. *et al.* Electrokinetically controlled real-time polymerase chain reaction in microchannel using Joule heating effect. *Anal. Chim. Acta* **557**, 146–151 (2006).
- Woolley, A. T. *et al.* Functional integration of PCR amplification and capillary electrophoresis in a microfabricated DNA analysis device. *Anal. Chem.* **68**, 4081–4086 (1996).
- Krishnan, M., Ugaz, V. M. & Burns, M. A. PCR in a Rayleigh–Benard convection cell. *Science* **298**, 793 (2003).

78. Kopp, M. U., deMello, A. J. & Manz, A. Chemical amplification: continuous-flow PCR on a chip. *Science* **280**, 1046–1048 (1998).
79. Obeid, P. J., Christopoulos, T. K., Crabtree, H. J. & Backhouse, C. J. Microfabricated device for DNA and RNA amplification by continuous-flow polymerase chain reaction and reverse transcription-polymerase chain reaction with cycle number selection. *Anal. Chem.* **75**, 288–295 (2003).
80. Marcus, J. S., Anderson, W. F. & Quake, S. R. Parallel picoliter RT-PCR assays using microfluidics. *Anal. Chem.* **78**, 956–958 (2006).
81. Liu, R. H., Yang, J., Lenigk, R., Bonanno, J. & Grodzinski, P. Self-contained, fully integrated biochip for sample preparation, polymerase chain reaction amplification, and DNA microarray detection. *Anal. Chem.* **76**, 1824–1831 (2004).
82. Lagally, E. T. *et al.* Integrated portable genetic analysis microsystem for pathogen/infectious disease detection. *Anal. Chem.* **76**, 3162–3170 (2004).
83. Blazej, R. G., Kumaresan, P. & Mathies, R. A. Microfabricated bioprocessor for integrated nanoliter-scale Sanger DNA sequencing. *Proc. Natl Acad. Sci. USA* **103**, 7240–7245 (2006).
84. Schwarz, M. A. & Hauser, P. C. Recent developments in detection methods for microfabricated analytical devices. *Lab Chip* **1**, 1–6 (2001).
85. Dittrich, P. S. & Manz, A. Single-molecule fluorescence detection in microfluidic channels — the holy Grail in μ TAS? *Anal. Bioanal. Chem.* **382**, 1771–1782 (2005).
86. Dertinger, S. K. W., Chiu, D. T., Jeon, N. L. & Whitesides, G. M. Generation of gradients having complex shapes using microfluidic networks. *Anal. Chem.* **73**, 1240–1246 (2001).
87. Jeon, N. L. *et al.* Neutrophil chemotaxis in linear and complex gradients of interleukin-8 formed in a microfabricated device. *Nature Biotechnol.* **20**, 826–830 (2002).
88. Mao, H., Yang, T. & Cremer, P. S. A microfluidic device with a linear temperature gradient for parallel and combinatorial measurements. *J. Am. Chem. Soc.* **124**, 4432–4435 (2002).
89. Cabrera, C. R., Finlayson, B. & Yager, P. Formation of natural pH gradients in a microfluidic device under flow conditions: model and experimental validation. *Anal. Chem.* **73**, 658–666 (2001).
90. Ratner, D. M. *et al.* Microreactor-based reaction optimization in organic chemistry — glycosylation as a challenge. *Chem. Commun.* 578–580 (2005).
91. Leung, S.-A., Winkle, R. F., Wootton, R. C. R. & deMello, A. J. A method for rapid reaction optimisation in continuous-flow microfluidic reactors using online Raman spectroscopic detection. *Analyst* **130**, 46–52 (2005).
92. Hatakeyama, T., Chen, D. L. & Ismagilov, R. F. Microgram-scale testing of reaction conditions in solution using nanoliter plugs in microfluidics with detection by MALDI-MS. *J. Am. Chem. Soc.* **128**, 2518–2519 (2006).
93. Chambers, R. D. *et al.* Elemental fluorine Part 16. Versatile thin-film gas-liquid multi-channel microreactors for effective scale-out. *Lab Chip* **5**, 191–198 (2005).
94. Lee, C. C. *et al.* Multistep synthesis of a radiolabeled imaging probe using integrated microfluidics. *Science* **16**, 1793–1796 (2005).
95. Kikutani, Y. *et al.* Pile-up glass microreactor. *Lab Chip* **2**, 193–196 (2002).
96. Pennemann, H., Watts, P., Haswell, S. J., Hessel, V. & Lowe, H. Benchmarking of microreactor applications. *Org. Proc. Res. Dev.* **8**, 422–439 (2004).

Acknowledgements I would like to thank C. deMello and T. deMello for help in preparing the manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence should be addressed to the author (a.demello@imperial.ac.uk).

Cells on chips

Jamil El-Ali¹, Peter K. Sorger² & Klavs F. Jensen¹

Microsystems create new opportunities for the spatial and temporal control of cell growth and stimuli by combining surfaces that mimic complex chemistries and geometries of the extracellular matrix with microfluidic channels that regulate transport of fluids and soluble factors. Further integration with bioanalytic microsystems results in multifunctional platforms for basic biological insights into cells and tissues, as well as for cell-based sensors with biochemical, biomedical and environmental functions. Highly integrated microdevices show great promise for basic biomedical and pharmaceutical research, and robust and portable point-of-care devices could be used in clinical settings, in both the developed and the developing world.

In their normal environment, cells are subject to multiple cues that vary in time and space, including gradients of cytokines and secreted proteins from neighbouring cells, biochemical and mechanical interactions with the extracellular matrix (ECM), and direct cell–cell contacts (Box 1). Microfabricated systems can present cells with these cues in a controllable and reproducible fashion that cannot easily be achieved by standard tissue culture, and can be used to link cell culture with integrated analytical devices that can probe the biochemical processes that govern cell behaviour. Some cell-based microsystems simply represent miniaturized versions of conventional laboratory techniques, whereas others exploit the advantages of small length scales and low Reynolds numbers¹, such as favourable scaling of electrical fields and the ability to create well-controlled laminar flows. In this Review, we discuss the application of microtechnology to cell biology and describe methods for cell culture, regulation of extracellular cues, cell fractionations and biochemical analysis on a micrometre scale (Fig. 1). Emphasis is placed on microsystems aimed at gaining biological insight, as well as on efforts to realize increasing cell-handling integration and biochemical analysis levels on chips.

We believe these devices will become increasingly implemented in applied and basic biomedical research, mainly because soft lithography² has put microfluidics within the reach of biology-focused academic laboratories. Elastomeric materials used in soft lithography, typically

poly(dimethylsiloxane) (PDMS), are relatively easy to fabricate, and are compatible with most biological assays. Devices that are based on micro-fabrication of silicon and glass require access to advanced cleanroom facilities similar to those used for microelectronics. This typically involves higher cost, but has unique advantages for specialized applications, such as electrophoresis in glass devices.

Much cell-based microsystem research takes place under a ‘lab-on-a-chip’ or ‘micro-total-analysis-system’ (μTAS) framework that seeks to create microsystems incorporating several steps of an assay into a single system^{3–5}. Integrated microfluidic devices perform rapid and reproducible measurements on small sample volumes while eliminating the need for labour-intensive and potentially error-prone laboratory manipulations. Thus, microfluidics allows experiments to be carried out that cannot be performed simply by miniaturizing and mechanizing conventional laboratory procedures using robotics and microplates. For example, in cell-based studies, the transition from 384- to 1,536-well plates is proving challenging, largely because edge effects and uncontrolled evaporation from very small wells result in poorly defined culture conditions. Conventional handling of very small fluidic volumes is difficult, and subject to both variability and high fixed losses. The fabrication of many copies of an analytic device, small reagent volumes, and diminished labour associated with use of automated microfabricated devices

Box 1 | Cell physiology, phenotype and fate are regulated by cell-autonomous processes and extracellular signalling molecules

Soluble signalling molecules include hormones, cytokines and growth factors produced by local or distant cells (giving rise to paracrine and endocrine signals, respectively), and even by the receiving cell itself (autocrine signals). Insoluble signalling molecules include components of the ECM and membrane-bound proteins on neighbouring cells. Cells sense most extracellular signals (including proteins, peptides and carbohydrates) via transmembrane receptors that activate complex biochemical cascades of kinases, proteases, adaptor proteins, transcription factors and so on, which together act to regulate cell physiology. At the same time, cells alter their surrounding environment by making or destroying ECM or soluble factors and by exerting mechanical force^{99,100}.

In animals, cells typically reside in environments with very specific three-dimensional (3D) features. Cells are sensitive to the presence of neighbouring cells of similar or different type and often make long-lasting mechanical and biochemical connections to them. In columnar epithelia, for example, identical cells lined up side by side assemble junctions with neighbours to form continuous impermeable sheets. These sheets are polarized such that cells interact with their surroundings in very different

ways on the luminal and basolateral surfaces. In addition, epithelia usually establish a close relationship with specific types of immune cell. In these epithelia, both homotypic and heterotypic cell–cell interactions are essential to maintain cell and tissue function.

Most *in vitro* experiments with adherent human cells are performed in two-dimensional (2D) cultures in which cells are plated onto plastic surfaces treated to stimulate cell binding. Depending on their type, cells either grow directly on the plastic, secrete ECM components that coat the plastic to facilitate cell adhesion, or require pre-coating of the plastic with ECM. Standard 2D culture conditions are poor mimics of the cellular environment in an animal: soluble growth factors are present at abnormally high concentrations, 3D cues are largely absent, oxygen tension is too high and cell–cell interactions are rarely organized. Attempts have been made to overcome these problems using organ culture and various laboratory-scale bioreactors, but microsystems provide a much more effective means of controlling cell environment *in vitro*. Particularly promising are various artificial organ systems in which multiple cell types are grown together under conditions that mimic normal 3D environmental and circulatory cues.

¹Department of Chemical Engineering, and ²Department of Biology and Biological Engineering, Center for Cell Decision Processes, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

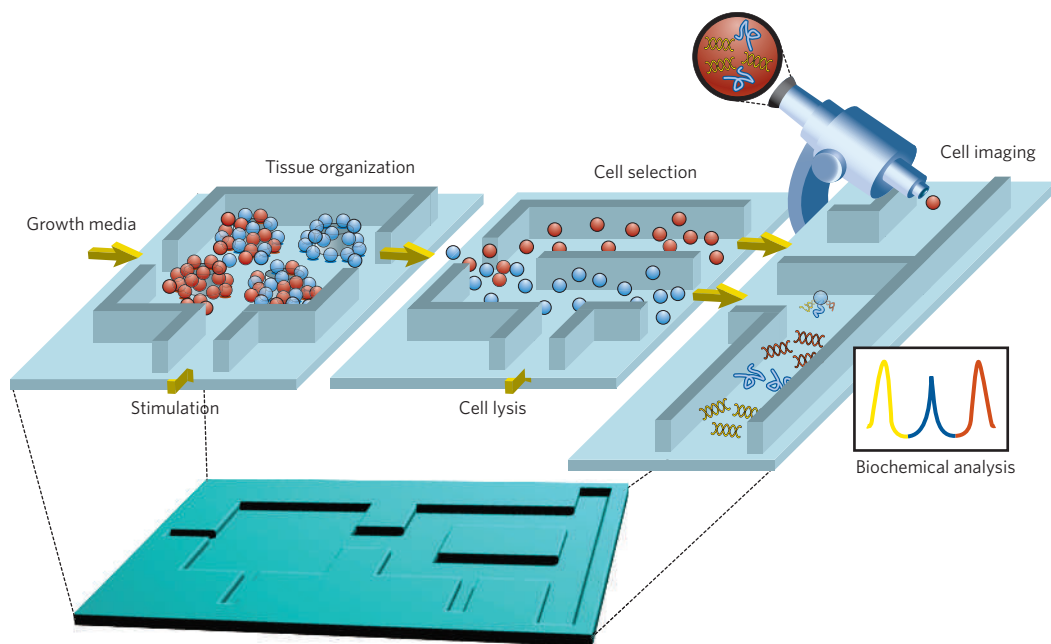


Figure 1 | Tissue organization, culture and analysis in microsystems. Advanced tissue organization and culture can be performed in microsystems by integrating homogeneous and heterogeneous cell ensembles, 3D scaffolds to guide cell growth, and microfluidic systems for transport of nutrients and other soluble factors. Soluble factors — for example, cytokines for cell stimulation — can be presented to the cells in precisely defined spatial and temporal patterns using integrated microfluidic systems. Microsystems technology can also fractionate heterogeneous cell populations into homogeneous populations, including single-cell selection, so different cell types can be analysed separately. Microsystems can incorporate numerous techniques for the analysis of the biochemical reactions in cells, including image-based analysis and techniques for gene and protein analysis of cell lysates. This makes microtechnology an excellent tool in cell-based applications and in the fundamental study of cell biology. As indicated by the yellow arrows, the different microfluidic components can be connected with each other to form an integrated system, realizing multiple functionalities on a single chip. However, this integration is challenging with respect to fluidic and sample matching between the different components, not least because of the difficulty in simultaneously packaging fluidic, optical, electronic and biological components into a single system.

should make them highly cost effective. Moreover, the small footprint and low power consumption of integrated systems creates opportunities for portable, point-of-care devices that can perform analyses hitherto possible only in the research or clinical laboratory. Devices such as these with sophisticated diagnostic capabilities are likely to become important in the personalization of medical care.

Many of the promises of μ TAS have yet to be realized: integration and packaging of several functionalities into a single system is proving to be a complex task (Fig. 1), and many cell-based microsystems available today are still in the proof-of-concept phase. Typical unit operations (for example, growth, treatment, selection, lysis, separation and analysis) have been demonstrated (Fig. 2), but robust approaches to fabrication, integration and packaging (such as communication with the macroenvironment) remain major areas of research.

Microfabricated cell cultures

Culturing cells *in vitro* is one of the cornerstones of modern biology. Nevertheless, even for intensively studied tissues, many of the factors that induce or stabilize differentiated phenotypes are poorly understood and difficult to mimic *in vitro*⁶. One approach to increase control over cell–cell and soluble cues typical of *in vivo* cell environments is to combine microfabrication of 3D ECM structures and microfluidic networks that transport soluble factors such as nutrients and oxygen. Microfluidics has the additional advantage of being capable of creating mechanical strain, through shear, in the physiological range.

Cells and the extracellular matrix

Microfabrication integrating micropatterning techniques with advanced surface chemistry makes it possible to reproducibly engineer cell micro-environment at cellular resolution. A large variety of surface-patterning techniques are available, including standard photolithography lift-off

techniques, photoreactive chemistry and, increasingly, techniques based on soft lithography (microcontact printing and fluidic patterning)⁷. Surface patterning of micrometre-sized features allows micrometre-scale control over cell–ECM interactions and can be used to generate ensembles of cells with defined geometry. Lamination, moulding and photo-polymerization techniques all allow fabrication of 3D scaffolds with feature sizes in the lower micrometre range, including microstructured scaffolds made of biodegradable materials⁸.

The precise control of the cellular environment that has been made possible by microtechnology provides new opportunities for understanding biochemical and mechanical processes responsible for changes in behaviour such as the effects of cell shape on the anchorage-dependence of cell growth^{9,10}. For example, by altering the spacing of a grid of cell-adhesive islands it is possible to control the extent of cell spreading, while keeping the cell–ECM contact area constant¹⁰ (Fig. 3a). Human capillary endothelial cells confined to closely spaced islands undergo apoptosis, whereas cells that can spread freely survive and proliferate normally¹⁰. Adhesive ECM patches can also be designed so that the locations of focal adhesions (integrin-mediated links between the ECM and actin cytoskeleton) result in the same overall cell shape, but with a different underlying cytoskeletal organization (Fig. 3b). By allowing cells to spread and proliferate on these adhesive patches the orientation of the cell division axis can be controlled¹¹. Similar regulation of the division axis by the ECM is likely to be important for tissue morphogenesis and other developmental processes.

The force exerted on the ECM by cells can be measured in several ways. A particularly powerful method involves measuring the deflection of arrays of micrometre-sized vertical elastomer posts (Fig. 3c). When tested with smooth muscle cells, forces acting in the plane of the substrate are in the range of 100 nN, and appear to scale with the area covered by focal adhesions¹². Compared with conventional methods that rely on substrate distortion, the elastomer-post technique has

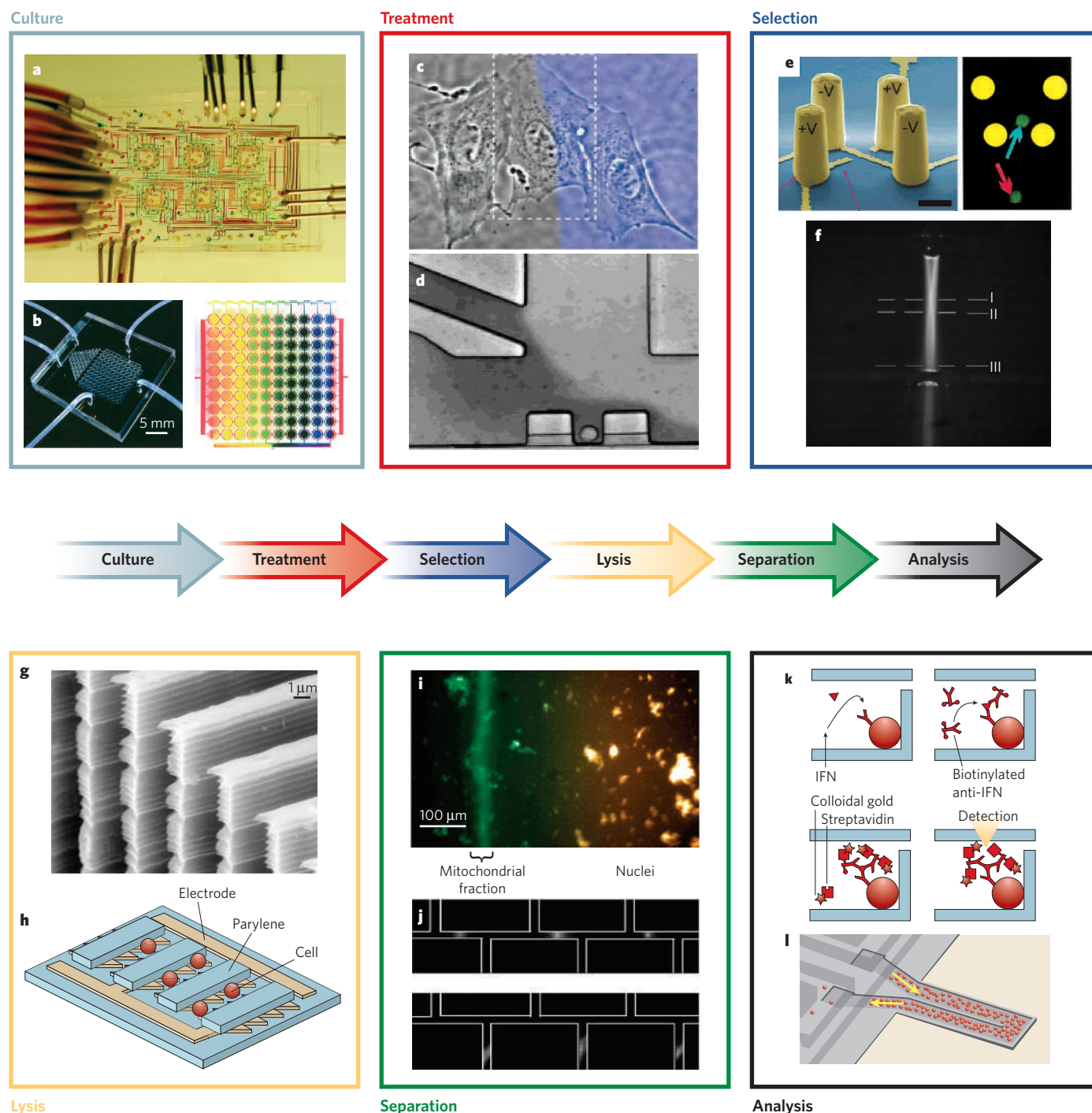


Figure 2 | Microsystems enabling cell-based assays from cell culture to biochemical analysis. A collection of microsystems enabling cell-based assays, covering all the steps from cell culture, through selection and treatment, to biochemical analysis. **a**, Image showing six bioreactors that can operate in parallel on a single chip. Each reactor can be used to monitor the growth of extremely small numbers of cells. (Image reproduced, with permission, from ref. 20.) **b**, Microfluidic cell-culture array with integrated concentration gradient generator (left). Image of concentration gradient across ten columns when loaded with blue and yellow dye. (Image reproduced, with permission, from ref. 33.) **c**, Two different laminar streams exposing two sides of a single cell to different conditions³⁴. **d**, Perfusion over a single hydrodynamically trapped cell. Switching of the perfused media can occur in ~100 ms. (Image reproduced, with permission, from ref. 38.) **e**, Single-cell dielectrophoresis (DEP) trap, consisting of four electroplated electrodes (left). Fluorescent image of a trapped cell (indicated by blue arrow; right). The cell has been loaded with calcein through the microfluidic system. (Image reproduced, with permission, from ref. 46.) **f**, Fluorescent image of light path at the detection zone in a micro flow cytometer with integrated waveguides and lenses. (Image reproduced, with permission, from ref. 53.) **g**, Scanning electron micrograph of a mechanical lysis device with sharp knife-like protrusions. (Image reproduced, with permission, from ref. 55.) **h**, Schematic of electrical lysis device with integrated microelectrodes. (Image reproduced, with permission, from ref. 56.) **i**, Isoelectric focusing of cell organelles from whole-cell lysate. The mitochondria focuses in a band at pI between 4 and 5. (Image reproduced, with permission, from ref. 62.) **j**, Two-dimensional separation of four model proteins. Isoelectric focusing (top) followed by SDS gel electrophoresis. (Image reproduced, with permission, from ref. 64.) **k**, Schematic of immunoassay performed using microbeads as solid support in a microfluidic system. (Image adapted, with permission, from ref. 69.) **l**, Schematic of a hollow cantilever-based mass sensor for analyte detection. (Image adapted, with permission, from ref. 74.)

the advantage of greater accuracy and manipulability: the mechanical properties of a surface can be varied by changing post geometries without altering surface chemistry¹².

Liver-cell culture

In vitro culture of liver cells has received particular attention in biotechnology as many drugs fail in clinical studies either because they damage the liver directly or because liver metabolites are toxic¹³. The study of hepatotoxicity would be greatly facilitated by the availability of *in vitro* culture systems that mimic real liver conditions. However, the development of liver-cell cultures as biosensors for drug toxicity faces challenges because of the difficulty in maintaining the differentiated phenotypes.

In the liver, hepatocytes are found in a complex 3D environment in which nutrients, soluble factors and oxygen are transported through blood capillaries and bile canaliculi. Using silicon as a substrate, perfused 3D liver reactors have been fabricated on arrays of 300- μ m-wide channels (capillaries) that comprise a scaffold for the ECM¹⁴ (Fig. 3d). Seeding hepatocytes with pre-aggregated multicellular spheroids in the 3D reactor generates cultures that are viable for a long time period (~3 weeks) and that exhibit a stable differentiated phenotype. Cells in 3D liver cultures also have cell–cell contacts, such as tight junctions and desmosomes, that resemble those found in tissues *in vivo*^{13,14}.

It has been observed that co-culture of hepatocytes with other cell types, including liver epithelial cells and Kupffer cells, prolongs the survival of cultured hepatocytes and helps maintain liver-specific properties such as albumin secretion¹⁵. Using a micropatterned 2D co-culture system, it has also been shown that liver-specific functions increase with heterotypic cell–cell interactions. Only hepatocytes close to the heterotypic interface maintain their differentiated phenotypes in longer-term culture⁶ (Fig. 3e). Relative to conventional co-culture, in which seeding densities of two cell types are varied on a planar surface, micropatterning techniques afford greatly improved control of homo- and heterotypic cell–cell interactions¹⁶. The ability to culture cells such as liver cells *in vitro* and to demonstrate protein and gene expression levels similar to those found in tissue suggests that microfabricated cultures could have applications in toxicology and could also serve as model systems for *in vitro* analogues of organ tissue.

Bone

Bone loss after menopause, long periods of inactivity or life in a microgravity environment poses a serious medical problem. Bone is a tissue in which shear stress and mechanical loading are important. Mechanical interactions are necessary for maintaining cultured osteoblastic cells in a state suitable for bone engineering. In standard 2D culture, shear stresses as low as 10 μ Pa enhance differentiation of osteoblasts as measured by alkaline phosphatase activity and fibronectin expression¹⁷. Microtechnology provides an opportunity to build 3D scaffold and fluidic networks that mimic the natural 3D environment of bone. This includes the use of fluidics to deliver soluble factors to the cells and to impose shear stress at the physiological level. In a 3D network of microstructured channels, alkaline phosphatase activity in osteoblast cells is enhanced threefold under static conditions (corresponding to a structural effect) and 7.5-fold under low flow conditions (representing a combined structural and shear effect) relative to 2D static cultures¹⁸.

Microorganisms

Microsystems also have applications as tools to screen and optimize conditions for yeast and *Escherichia coli* fermentation and growth during bioprocessing. Microfluidic bioreactors have been miniaturized to create nanolitre growth chambers in which extremely small cell populations can be monitored^{19,20} (Fig. 2a). The use of small reactor volumes and multiple independent cell populations helps to decrease problems associated with genetic variation and makes it possible to assess many different conditions in parallel. The ability to integrate optical sensors in growth chambers also makes it possible to monitor key process variables such as pH, dissolved oxygen and biomass²¹. Data on these variables could be combined with gene expression analysis and metabolic studies to rapidly prototype and then scale up conditions for industrial bioprocessing²².

Cell stimulation and selection

The control of cellular microenvironments via microfluidic systems potentially represents a valuable tool for fundamental studies of cell biology. Biological insight into the pathways that control cell phenotype and behaviour can be gained by monitoring cellular responses to controlled perturbations in the extracellular environment. A wide range of microsystems are therefore emerging with the express aim of facilitating the basic study of biochemical pathways, cell-fate decisions and tissue morphogenesis. In the next two sections, we provide examples of some techniques being applied to cell-based assays (Fig. 2). Readers are also referred to more detailed reviews elsewhere^{23–29}.

Cell stimulation of adherent cells

Controlled perturbation of the cellular environment in time and space for adherent cells in microfluidic devices can be accomplished by controlling flow over the cells. As mentioned above, fluid flow not only transports soluble factors, but also exerts mechanical force through shear^{14,18}. The diffusive mixing properties of laminar flow created by microfluidics can also be used to create complex concentration gradients not achievable on a macroscale³⁰. These gradients allow several conditions to be probed simultaneously while also mirroring conditions found *in vivo*. For example, repeated combinations of flow-stream lamination and splitting can create complex concentration gradients that promote cell chemotaxis³⁰ — the migration of cells in response to a stimulus. Some bias in the cell migration due to shear is observed in systems with high flow rates³¹. Linear gradients of external factors can also be created in static (convection-free) microfluidic systems without perturbing the existing distribution of secreted molecules, thereby preserving autocrine and paracrine signalling³². For cell culture applications, gradient generation permits many growth conditions to be analysed in a combinatorial fashion³³ (Fig. 2b).

In systems with fast flow or large molecules (small diffusion coefficients), diffusion is often too slow for any appreciable mixing between fluidic streams. Although slow diffusion poses a complication when mixing is desired, slow diffusive mixing creates opportunities for varying the liquid-phase environment over distances comparable to the size of cells. For example, laminar flow has been used to expose two halves of an endothelial cell to different mitochondrial dyes, making it possible to observe the movement of organelles from one side of a cell to the other³⁴ (Fig. 2c). A similar approach, based on temperature steps rather than a chemical gradient, has recently been used to study the effects of temperature perturbations on embryonic development in *Drosophila*³⁵.

Cell stimulation in suspension

Cells in suspension are usually transported with flow. However, cells can be physically retained in the devices by filters or traps. By combining concentration gradients and flow of cells with hydrodynamic traps that retain cells in a fixed position, microfluidics has been used to monitor ATP-dependent calcium uptake in HL-60 cells³⁶. Hydrodynamic traps placed on either side of a narrow microfluidic channel can also be used to capture cells so that well defined cell–cell contacts form³⁷. Trapping single cells near a two-channel reagent-delivery system can also enable very rapid (within ~100 ms) fluidic switching between a buffer and a reagent stream at the cell position. This makes it possible to monitor very fast cellular responses, such as calcium flux in ionomycin-stimulated Jurkat cells³⁸ (Fig. 2d).

If cells are allowed to flow through a device with the bulk liquid, they will spend a characteristic time in each fluidic compartment. Soluble cues can then be applied to cells for varying time periods by mixing in reagents in successive fluidic compartments. For example, by using a flow segmentation technique to enhance mixing, fast transient mitogen-activated protein kinase responses of α -CD3-stimulated Jurkat cells have been monitored with excellent reproducibility and temporal control³⁹. Flow segmentation enhances mixing by creating a small recirculation within each segment⁴⁰. Analysis of stress markers shows that segmented microfluidic flow does not trigger significant cell stress responses³⁹.

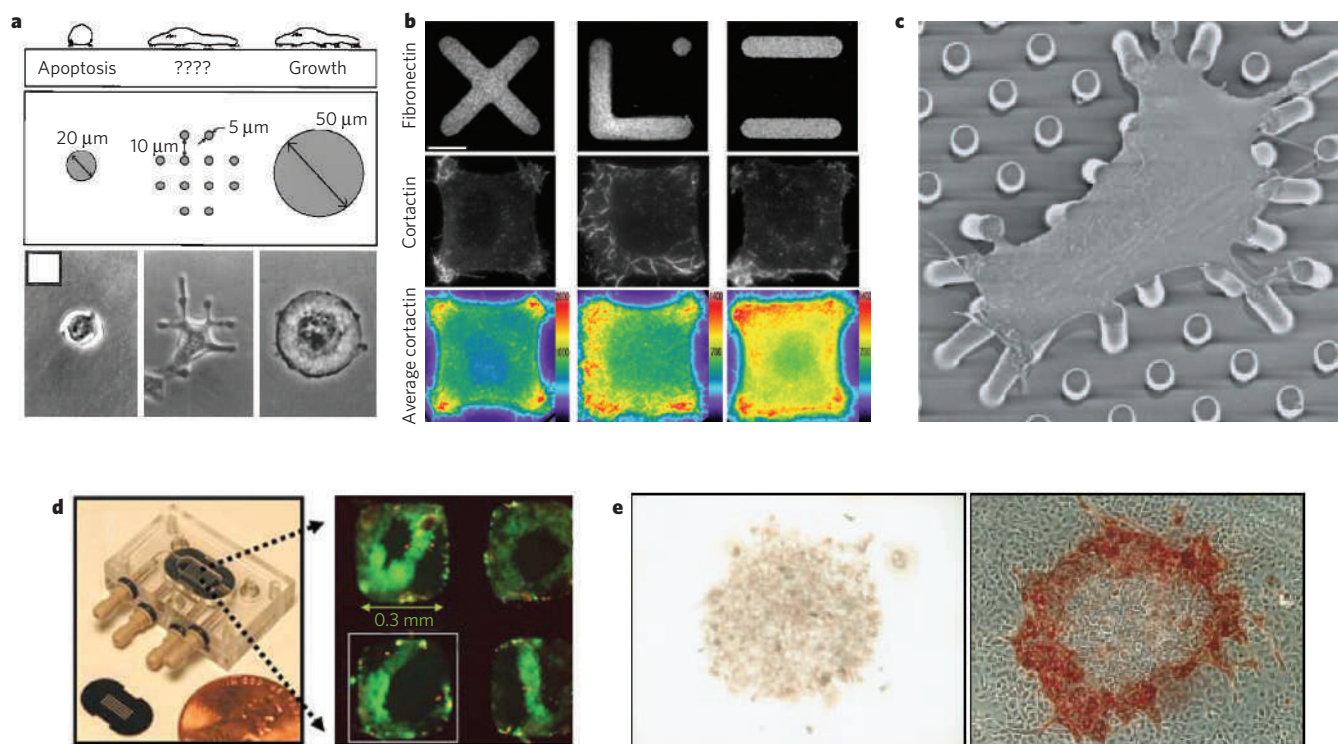


Figure 3 | Substrate patterning and tissue culture. **a**, Diagram of substrate patterns that can be used to control the area cells can spread over without varying the cell–ECM contact area. The corresponding images show that if cells are confined to a small area, they undergo apoptosis, whereas if they are allowed to spread over a larger area while keeping the same cell–ECM contact area, they remain viable. (Image reproduced, with permission, from ref. 10.) **b**, The membrane ruffles, revealed by the cortactin marker, are preferentially located where the cell membrane attaches to the fibronectin in the ECM¹¹. **c**, Cell cultured on an array of compliant micro-posts. The direction and magnitude of the deflection of the posts is a measure of the local force field. (Image reproduced, with permission, from ref. 12.) **d**, Assembled liver-cell microfluidic system with four ports for fluidic access. A viability stain shows that most cells in the scaffold are viable (green) and there are only few non-viable cells (red). (Image reproduced, with permission, from ref. 13.) **e**, Immunostaining of intracellular albumin in micropatterned hepatocyte cultures. Cells in the homoculture (left) have lost albumin after 6 days of culture. In the heteroculture (right), hepatocyte cells near the heterotypic interface retain albumin content at day 6, whereas cells away from this interface lose albumin content. (Image reproduced, with permission, from ref. 6.)

Cell sorting

Fluidic transport with selective trapping or diversion of suspended cells allows cell sorting to be integrated into microfluidic systems — a powerful capability when combined with the methods for cell stimulation. The ability to isolate homogeneous and concentrated cell populations from heterogeneous cell mixtures can also be important for obtaining accurate information about the underlying biochemistry of specific cell types in a mixture. Microtechnology makes it possible to isolate a few cells (or even single cells) from a large population of cells on the basis of physical and chemical properties such as electrical characteristics or fluorescent markers, ultimately allowing heterogeneities within seemingly homogeneous cell populations to be exposed⁴¹.

The ease of integrating electrodes along with the favourable scaling of electrical fields in microsystems creates opportunities for exploiting dielectrophoresis (DEP) to move, separate and position individual cells⁴². The DEP force arises from induced dipoles in cells exposed to a non-uniform electrical field. DEP depends on the intrinsic electrical properties of a cell, such as membrane capacitance and conductance, both of which change with cell type and even with cell activation⁴³. For example, by using DEP and microfluidics, MDA231 cancer cells have been separated from dilute blood by selective capture onto microelectrodes⁴⁴. Tagging cells with marker particles with different dielectric properties has allowed DEP sorting for rare cells at rates up to 10,000 cells s⁻¹ and with enrichment factors of more than 200 (ref. 45). With proper electrode design, DEP can be used to capture and manipulate single cells, and stimuli can be introduced via the fluidic system, for example, to monitor the kinetics of fluorescent dye (calcein) uptake in HL-60 cells⁴⁶ (Fig. 2e). Each trap can then be made electrically addressable for selective capture and release of cells for further analysis. The importance of separating cell populations before biochemical

analysis is exemplified by the DEP-based separation of U937 cells and peripheral blood mononuclear cells (PBMCs) into two homogeneous populations. Following cell stimulation, increases in the expression of cytokine genes can be detected in sorted populations of U937 cells, but the effect is almost completely masked in mixed-cell populations⁴⁷.

Fluorescently activated cell sorters (FACS) have also been miniaturized using integrated pneumatically activated pumps and valves that divert cells into a collection chamber on the basis of their fluorescent properties⁴⁸. Depending on cell concentrations and purity, cells can be sorted at rates of up to ~40 cells s⁻¹ with enrichment factors of ~90 and recovery yields of between 16 and 50%. Using optical forces instead of mechanical valves to switch the direction of cells permits a slightly higher throughput of ~100 cells s⁻¹, with recovery yields above 80% and enrichment factors of up to ~70 (ref. 49).

Cells can also be separated on the basis of the affinity of cell-surface receptors for proteins immobilized on the surfaces of the microfluidic channel⁵⁰. Transient adhesion between cells and appropriate surface ligands retard cell movement through the channel. This creates a chromatographic separation between two cell types on the basis of differences in their retardation. For example, by using selectin as an adhesion molecule, HL-60 cells can be separated from U937 cells, albeit with low resolution⁵⁰. A similar affinity-based capture technique has been used to capture T cells in microwells using anti-CD5 as a surface ligand⁵¹.

Although many microfluidic sorting systems achieve high levels of integration with electrodes, valves and even pumps, most microsystems rely on bulk optical elements such as lenses and microscopes external to the microdevice for optical control and detection. It is not uncommon to see a microchamber coupled to lenses and electronics that occupy an entire optical table. Miniaturization of free-space optical elements is not

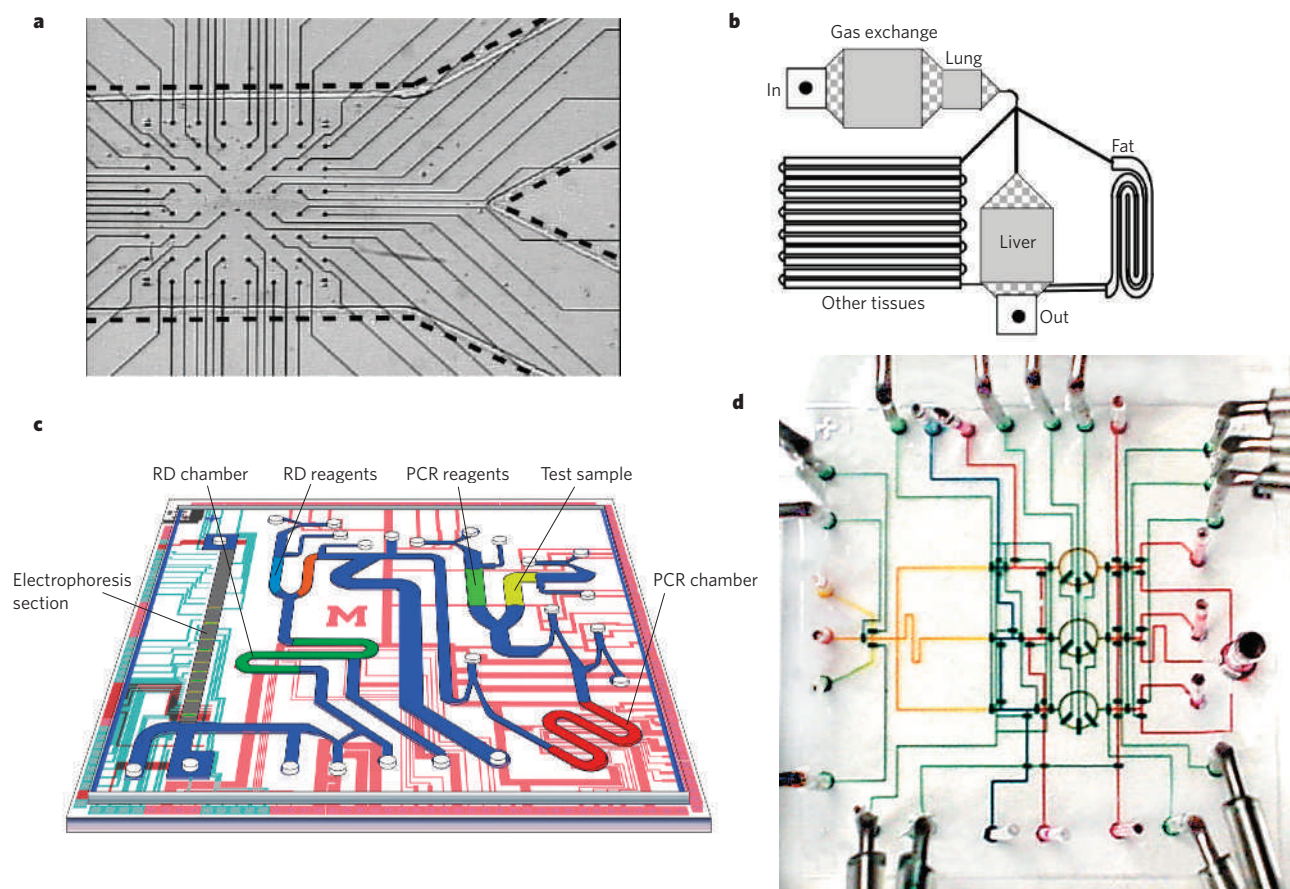


Figure 4 | Integrated cell analysis systems. **a**, Microelectrode array for recording neuronal activity integrated with a microfluidic channel. (Image reproduced, with permission, from ref. 81.) **b**, Cell culture analogue with four different interconnected tissue compartments. The lung and liver chambers contain cells, whereas the 'other tissues' and fat compartments have no cells but mimic the fluid residence-time distribution in tissues of rapid and slow perfusion, respectively. (Image adapted, with permission, from ref. 89.) **c**, Schematic of an integrated system for genetic analysis. The system can perform two independent serial biochemical reactions — polymerase chain reaction (PCR) amplification of a sample followed by restriction endonuclease digestion (RD). Analysis is performed by electrophoretic separation and fluorescent detection. (Image reproduced, with permission, from ref. 93.) **d**, DNA extraction and purification chip, with advanced integrated fluidic handling of cell samples as well as the necessary buffers and reagents⁹⁶.

easy, but efforts are underway to integrate optics with microsystems⁵². Recently, a microchip-based flow cytometer with integrated polymer waveguides and lenses was described⁵³ (Fig. 2f).

Most integrated microflow cytometers have yet to match conventional systems in performance ($>10,000$ cells s^{-1}), but they have smaller footprints, are of lower cost, and create opportunities for in-line integration with other analytic devices. Furthermore, with the laminar-flow conditions found in microsystems, the sorting of viable cells does not appear to perturb cell physiology appreciably^{47,49}.

Biochemical analysis of cell lysates

A significant research effort is devoted to the development of integrated tools for microscale biochemical analysis. Quantitative analysis of complex biochemical mixtures, such as cell lysates, remains challenging, and with many devices success has only been achieved with low-complexity samples. Nonetheless, almost every analytical tool available in a conventional biology lab has an equivalent microfabricated counterpart, and many of these have been nicely summarized in reviews^{23–25,27}. Protein analysis is generally more difficult than analysis of nucleic acids, because the physical and chemical properties of proteins are much more variable than those of either RNA or DNA. Moreover, unlike DNA and RNA, no methods exist for amplifying proteins. Relative abundance of proteins in cell lysates can vary by more than 10^5 , making sensitivity and dynamic range critical to any successful assay. The problems of low abundance and high complexity are generally handled in one of two ways: by linking sample preparation steps such as physiochemical separation and

concentration before analysis, or by using high selectivity in the analytical system, typically through affinity methods based on antibodies. In addition, it is important to apply surface coatings to limit non-specific surface absorption, which results in significant loss of material.

Cell lysis

The lysis of cells in the laboratory is generally accomplished either chemically, with detergents, or mechanically, by membrane rupture. Microfluidic devices can incorporate either method. Chemical lysis with Triton X-100 (ref. 39) and denaturation with sodium dodecyl sulphate (SDS)⁵⁴ have both proved effective on a microscale. The main advantage of chemical lysis is that subsequent assays can be performed in buffers previously optimized for conventional biological studies. Mechanical cell lysis is an alternative method if detergents interfere with downstream analysis. For example, microscale shear lysis similar to conventional macroscopic techniques can be achieved by forcing HL-60 cells through nanoscale barbs⁵⁵ (Fig. 2g). Although lysis — as monitored by release of an intracellular dye — is almost complete, aggregation of the cellular debris results in such inefficient protein extraction that only ~5% of total protein is accessible for subsequent analysis. The use of electroporation as an alternative to mechanical lysis has been motivated by its ability to achieve high local fields using integrated microelectrodes^{56,57} (Fig. 2h). Electrical lysis can be rapid, with disruption times as low as 33 ms — about eight times faster than lysis by SDS⁵⁸. By carefully controlling the strength of the electrical field, microfabricated electroporation devices can also reversibly destabilize the cell membrane for gene transfection applications⁵⁹.

Sample preparation

Protein and DNA fractionation has been achieved using microfabricated sieving systems with nanometre-sized filters^{60,61}. However, microsystems that rely on electrokinetic separation techniques, such as capillary electrophoresis (CE), gel electrophoresis, electrochromatography or isoelectric focusing (IEF)^{23,24,27} are more commonly used. These microsystems can typically achieve faster separation with smaller volumes than their conventional counterparts. The use of electrokinetic techniques also extends to the separation of organelles. Mitochondria from HeLa cells undergoing apoptosis have been fractionated from each other and from nuclei using IEF⁶² (Fig. 2i). Increased resolution in electrokinetic protein separation on the microscale can be achieved by using a 2D separations system similar to those used on the laboratory scale — for example, a combination of electrochromatography with CE⁶³, or IEF with gel electrophoresis⁶⁴ (Fig. 2j). Electrokinetic techniques can also be used to concentrate proteins and peptides significantly in a sample⁶⁵.

The ability to amplify DNA and RNA using PCR and reverse transcription PCR (RT-PCR) is a very powerful tool that allows sensitive detection and quantification of nucleic acids. Microdevices for PCR amplification integrate electrodes for heating and temperature sensing used to control the thermal reactions of the PCR cycle. The low thermal mass of microsystems results in very fast temperature cycling, with heating and cooling rates in excess of 35 °C s⁻¹ (ref. 66). DNA amplification can also be directly integrated with electrophoretic separation techniques for complete sample pretreatment systems⁶⁷.

Analytical techniques

A large number of techniques exist for the quantification of cell lysates. A major advantage of microsystems containing integrated electrophoretic separation technology is that detection and quantification are possible by absorption techniques or by using relatively simple fluorescent markers. After separation, the sample simply passes by an optical detector. Microsystems with electrophoretic separation are also increasingly being coupled to mass spectrometers for protein identification and the analysis of post-translational modifications^{23,24,68}.

One quantitative method for protein analysis and detection that is widely used in microsystems and does not require prior fractionation of even complex biological samples is antibody capture. Antibody-based techniques rely on the high selectivity and affinity of antibody–antigen binding to achieve specificity in analysis^{24,27}. Integrated microfluidic systems can coat either solid supports⁶⁹ (Fig. 2k) or channel surfaces⁷⁰ with capture antibodies, and subsequently introduce the analyte and secondary detection antibodies, if necessary. Similar microsystems have been developed for DNA hybridization, with some devices able to carry out a hybridization assay on 1 µl of sample in less than 10 min (ref. 71).

Label-free detection represents a potentially powerful alternative to fluorescent and luminescent detection. Generally applicable methods include detection of changes in mass and electrical properties due to binding on antibody-functionalized sensor surfaces. Field-effect sensors have been realized with carbon nanotubes⁷² and silicon nanowires⁷³ as the active sensing elements. Recently described cantilever technologies look promising as mass-based sensor techniques^{74,75} (Fig. 2l). The extensive area available for affinity capture and the sensitivity of these systems should allow them to detect femtomoles of protein.

The ability of analytical microfluidic systems to reproducibly handle small samples and high throughput, combined with the integration of sample preparation and separation steps, has resulted in these systems finding considerable commercial applications in the fields of genomics and proteomics⁷⁶.

Applications

Cell-based microdevices, including biosensors, are increasingly being used in drug discovery, genetic analysis and single-cell analysis. This section describes some recent examples, although many others can be found in the literature, and new devices and applications continuously emerge.

Biosensors

Cell-based biosensors monitor physiological changes in reporter cells exposed to biological or industrial samples containing pathogens, pollutants, biomolecules or drugs^{77,78}. The readout can be optical (for example, fluorescent, luminescent or colorimetric) or electrical (for instance, measuring changes in impedance or electrical potential)⁷⁹. Some biosensors detect simple phenotypic responses, such as life versus death.

The chemical-dependent electrophysiological activity of certain cell types, such as neurons and cardiac cells, has spurred their use in chip-based biosensors. Changes in electrical activity can be monitored by planar micro-electrode arrays⁸⁰, which are easily integrated into microfluidic devices and can be made with large numbers of measurement points per device⁸¹ (Fig. 4a). ECM surface patterning makes it possible to place neurons at precise points on the electrode arrays, and patterning can also be used to modulate the activity of the assembled neural networks *in vitro*⁸². Microfluidics can not only deliver soluble factors, but also present topographical cues that help to control neuronal connectivity⁸³. A portable, highly integrated cell-based biosensor system for the analysis of biochemical agents has been realized by integrating a complementary metal oxide semiconductor (CMOS) chip as digital interface with recording electrodes as well as with a temperature control system that includes heater electrodes to sustain the environmental requirements of the cells in the microfluidic culture chamber⁸⁴. The system has been tested with a 5-µM solution of a calcium channel blocker (nifedipine). Challenges still remain in using living cells as sensors, because variables such as cell density and cell interaction can significantly affect the sensor properties.

Drug screening

High-throughput measurement of ion-channel activity by patch clamping is of considerable interest in drug discovery as a tool to characterize therapeutic molecules. Microsystems that combine high throughput with small reagent volumes have led to commercial microscale patch-clamp devices⁸⁵. In these devices, ion-channel recording is typically achieved by placing cells on a micrometre-sized aperture in a membrane that separates two electrodes^{86,87}. By guiding cells onto apertures using microfluidic paths, it is possible to reduce the otherwise labour-intensive micromanipulations needed to locate cells at recording sites and to present the cell with successive stimuli⁸⁸. Obtaining the high-electrical-resistance seals necessary for high quality ion-channel recording (~10⁹ Ω) is technically challenging on both a macro- and a microscale, and microsystems have been more successful in meeting the throughput challenge.

One challenge in drug and toxicology screening is recreating the cell–cell interactions found in living organisms. The toxic effect of many drugs in a target tissue often depends on the metabolic activity of another tissue — in particular the liver. In such situations, tests of a drug on two tissues in isolation would not necessarily reveal any toxicity. This limitation has been addressed by constructing a microsystem using interconnected channels and chambers, each of which contains a different cell type mimicking the activity of a particular tissue⁸⁹ (Fig. 4b). The interconnected compartments for the liver, lungs and fat cells are designed to capture physiologically relevant features, such as residence times, of the circulation and interchange of metabolites in the body. However, challenges remain in maintaining differentiated phenotypes and matching fluidic conditions in different compartments⁸⁹.

Stem cells

The promise of stem cells for cell-based therapies in human disorders and tissue engineering has resulted in a growing interest in applying microtechnology to stem-cell culture. The controlled microenvironment of microfluidic platforms can be very useful in the study of stem cells⁹⁰. Manipulating the chemical environment of the culture in time and space allows the behaviour of stem cells, such as proliferation and differentiation, to be controlled.

A microfluidic stem-cell culture platform with a concentration gradient has been used to study the effect of growth-factor concentration on human neuronal-stem-cell behaviour⁹¹. The observed proliferation rate in the device was proportional to growth-factor concentration,

whereas differentiation (to astrocytes) was inversely proportional. In these studies, flow in the device minimized autocrine and paracrine signalling. However, it is also possible to set up a linear concentration gradient in a static microfluidic system, preserving autocrine and paracrine signals³². Recently, a microfluidic device for stem-cell culture with both logarithmic varying perfusion rates and concentration gradients has been developed, making it possible to explore a wide range of biological conditions (including effect of shear) simultaneously⁹². Future integration of advanced culturing techniques using heterotypic culture and 3D cues is likely to further increase the value of microdevices for stem-cell research.

Genetic analysis

The most developed analytical microsystems so far are those that measure DNA and RNA. These devices often rely on PCR and similar techniques for sample amplification, and include hybridization arrays, real-time probes or electrophoretic sizing for analysis. Pathogen and disease detection can benefit greatly from fast and cheap field-capable devices similar to the one recently developed for the detection of influenza⁹³ (Fig. 4c). This device integrates valves for precise fluidic handling, temperature control with integrated heaters for DNA amplification by PCR, and electrophoretic separation after restriction endonuclease digestion. The current cost of the device is estimated at US\$7 per chip, which could potentially drop below \$1 with further scaling down of device dimensions. A similar highly integrated but portable device is capable of bacterial pathogen detection after PCR amplification and electrophoretic separation in less than 10 min, with detection limit as low as 2–3 bacterial cells⁹⁴. Thermal cycling, sample purification and capillary electrophoresis have also been integrated in a device for nanolitre-scale DNA sequencing, allowing more than 500 continuous bases to be sequenced with 99% accuracy⁹⁵.

Single-cell analysis

In both conventional studies and microsystems, the analysis of single cells has typically been performed using image-based techniques and intracellular fluorescent probes (such as those that measure calcium flux³⁸). However, the ability of integrated microfluidics to accurately manipulate, handle and analyse very small volumes has opened up new opportunities for analysis of intracellular constituents. A microfluidic device with integrated pneumatic valves capable of isolating single cells and then lysing them using a chemical lysis buffer has been shown to be capable of extracting and recovering messenger RNA from a single cell⁹⁶ (Fig. 4d). A similar device that also integrates electrophoretic separation can analyse amino acids from the lysed contents of a single cell⁹⁷. Single-cell analysis by electrophoretic separation but with electrokinetic flow-driven cell loading, docking and lysis have also been demonstrated⁹⁸.

Outlook

Microfabricated devices have been developed to facilitate both applied and basic research into the biology of cells and tissues. However, many devices have so far only been tested with simple, low complexity samples, and examples of multi-step integration are only now emerging. In many cases in which actual biological specimens have been examined, it has been necessary to fractionate or otherwise process samples before introduction into the microsystem, although nucleic-acid analysis is one exception. Analysis of proteins in clinical samples, such as blood serum or whole-cell lysates, presents challenges. The realization of effective devices for pretreatment or fractionation of complex samples therefore remains a challenge to the practical application of integrated micro-analysis systems in protein chemistry. Integration and automation are important goals that also remain considerable hurdles. The rationales for integration include greater accuracy and reproducibility, smaller sample sizes, and higher throughput. Rather than monitoring only simple phenotypic changes, future integrated systems should be able to gather precise biochemical and mechanistic data from cells and tissues. A fully integrated liver toxicology chip, for example, might include 3D microculture that sustains the differentiated phenotypes

of multiple cell types, including hepatocytes, and fluidics to refresh the culture medium and apply biological cues or small molecules. On-line cell separators would facilitate the selection of specific cell populations that could then be delivered to a lysis chamber and, subsequently, to multiplexed on-line sandwich immunoassays using integrated optics or label-free detection. Such systems would require much less material than today's laboratory-scale methods, a huge advantage with primary cells and patient tissues.

The growing emphasis in molecular biology on single-cell analysis derives from increasing appreciation of phenotypic heterogeneity among cells in a population and of the scientific insight that derives from accurately assaying this heterogeneity. Physicochemical modelling of biological processes also demands single-cell data, or at least information about the distributions of key parameters. However, notable challenges remain in the detection of low-abundance proteins, which tend to adhere nonspecifically to surfaces (which are larger per unit volume in many microsystems than in laboratory-scale devices). Trade-offs are likely to exist between measuring more variables and using fewer cells. In our opinion, excessive emphasis on single-cell analysis, rather than on the use of microtechnology to link complex heterogeneous cultures, controlled perturbations and cell fractionation, is unwarranted (Fig. 1).

Although significant challenges face routine applications of 'cells on chips', tremendous advances have been realized over the past decade, and a future in which chips effectively compete with laboratory-scale technologies in the analysis of complex biological phenomena is clearly in sight. Highly integrated microdevices will find application in basic biomedical and pharmaceutical research, whereas robust and portable point-of-care devices will be used in clinical settings. ■

1. Stone, H. A., Stroock, A. D. & Ajdari, A. Engineering flows in small devices: microfluidics toward a lab-on-a-chip. *Annu. Rev. Fluid Mech.* **36**, 381–411 (2004).
2. Xia, Y. & Whitesides, G. M. Soft lithography. *Annu. Rev. Mater. Sci.* **28**, 153–184 (1998).
3. van den Berg, A. & Lammerink, T. S. J. in *Microsystem Technology in Chemistry and Life Science* (eds Manz, A. & Becker, H.) 21–49 (Springer, Berlin, 1998).
4. Manz, A., Graber, N. & Widmer, H. M. Miniaturized total chemical-analysis systems — a novel concept for chemical sensing. *Sens. Actuators B Chem.* **1**, 244–248 (1990).
5. Dittrich, P. S. & Manz, A. Lab-on-a-chip: microfluidics in drug discovery. *Nature Rev. Drug Discov.* **5**, 210–218 (2006).
6. Bhatia, S. N., Balis, U. J., Yarmush, M. L. & Toner, M. Effect of cell–cell interactions in preservation of cellular phenotype: cocultivation of hepatocytes and nonparenchymal cells. *FASEB J.* **13**, 1883–1900 (1999).
7. Folch, A. & Toner, M. Microengineering of cellular interactions. *Annu. Rev. Biomed. Eng.* **2**, 227–256 (2000).
8. Tsang, V. L. & Bhatia, S. N. Three-dimensional tissue fabrication. *Adv. Drug Deliv. Rev.* **56**, 1635–1647 (2004).
9. Britland, S., Clark, P., Connolly, P. & Moores, G. Micropatterned substratum adhesiveness — a model for morphogenetic cues controlling cell behavior. *Exp. Cell Res.* **198**, 124–129 (1992).
10. Chen, C. S., Mrksich, M., Huang, S., Whitesides, G. M. & Ingber, D. E. Geometric control of cell life and death. *Science* **276**, 1425–1428 (1997).
11. Théry, M. et al. The extracellular matrix guides the orientation of the cell division axis. *Nature Cell Biol.* **7**, 947–953 (2005).
12. Tan, J. L. et al. Cells lying on a bed of microneedles: an approach to isolate mechanical force. *Proc. Natl Acad. Sci. USA* **100**, 1484–1489 (2003).
13. Sivaraman, A. et al. A microscale in vitro physiological model of the liver: predictive screens for drug metabolism and enzyme induction. *Curr. Drug Metab.* **6**, 569–591 (2005).
14. Powers, M. J. et al. A microfabricated array bioreactor for perfused 3D liver culture. *Biotechnol. Bioeng.* **78**, 257–269 (2002).
15. Guguenguillou, C. et al. Maintenance and reversibility of active albumin secretion by adult-rat hepatocytes co-cultured with another liver epithelial-cell type. *Exp. Cell Res.* **143**, 47–54 (1983).
16. Bhatia, S. N., Yarmush, M. L. & Toner, M. Controlling cell interactions by micropatterning in co-cultures: hepatocytes and 3T3 fibroblasts. *J. Biomed. Mater. Res.* **34**, 189–199 (1997).
17. Liegibel, U. M. et al. Fluid shear of low magnitude increases growth and expression of TGF beta 1 and adhesion molecules in human bone cells in vitro. *Exp. Clin. Endocrinol. Diabetes* **112**, 356–363 (2004).
18. Leclerc, E. et al. Study of osteoblastic cells in a microfluidic environment. *Biomaterials* **27**, 586–595 (2006).
19. Groisman, A. et al. A microfluidic chemostat for experiments with bacterial and yeast cells. *Nature Methods* **2**, 685–689 (2005).
20. Balagadde, F. K., You, L. C., Hansen, C. L., Arnold, F. H. & Quake, S. R. Long-term monitoring of bacteria undergoing programmed population control in a microchemostat. *Science* **309**, 137–140 (2005).
21. Szita, N. et al. Development of a multiplexed microbioreactor system for high-throughput bioprocessing. *Lab Chip* **5**, 819–826 (2005).
22. Boccazzi, P. et al. Gene expression analysis of *Escherichia coli* grown in miniaturized bioreactor platforms for high-throughput analysis of growth and genomic data. *Appl. Microbiol. Biotechnol.* **68**, 518–532 (2005).

23. Vilkner, T., Janasek, D. & Manz, A. Micro total analysis systems. Recent developments. *Anal. Chem.* **76**, 3373–3385 (2004).
24. Lion, N. *et al.* Microfluidic systems in proteomics. *Electrophoresis* **24**, 3533–3562 (2003).
25. Auroux, P. A., Koc, Y., deMello, A., Manz, A. & Day, P. J. R. Miniaturised nucleic acid analysis. *Lab Chip* **4**, 534–546 (2004).
26. Andersson, H. & van den Berg, A. Microfluidic devices for cellomics: a review. *Sens. Actuators B Chem.* **92**, 315–325 (2003).
27. Verpoorte, E. Microfluidic chips for clinical and forensic analysis. *Electrophoresis* **23**, 677–712 (2002).
28. Breslauer, D. N., Lee, P. J. & Lee, L. P. Microfluidics-based systems biology. *Mol. Biosys.* **2**, 97–112 (2006).
29. Helmke, B. P. & Minerick, A. R. Designing a nano-interface in a microfluidic chip to probe living cells: challenges and perspectives. *Proc. Natl Acad. Sci. USA* **103**, 6419–6424 (2006).
30. Jeon, N. L. *et al.* Neutrophil chemotaxis in linear and complex gradients of interleukin-8 formed in a microfabricated device. *Nature Biotechnol.* **20**, 826–830 (2002).
31. Walker, G. M. *et al.* Effects of flow and diffusion on chemotaxis studies in a microfabricated gradient generator. *Lab Chip* **5**, 611–618 (2005).
32. Abhyankar, V. V., Lokuta, M. A., Huttenlocher, A. & Beebe, D. J. Characterization of a membrane-based gradient generator for use in cell-signaling studies. *Lab Chip* **6**, 389–393 (2006).
33. Hung, P. J., Lee, P. J., Sabounchi, P., Lin, R. & Lee, L. P. Continuous perfusion microfluidic cell culture array for high-throughput cell-based assays. *Biotechnol. Bioeng.* **89**, 1–8 (2005).
34. Takayama, S. *et al.* Laminar flows — subcellular positioning of small molecules. *Nature* **411**, 1016 (2001).
35. Lucchetta, E. M., Lee, J. H., Fu, L. A., Patel, N. H. & Ismagilov, R. F. Dynamics of *Drosophila* embryonic patterning network perturbed in space and time using microfluidics. *Nature* **434**, 1134–1138 (2005).
36. Yang, M. S., Li, C. W. & Yang, J. Cell docking and on-chip monitoring of cellular reactions with a controlled concentration gradient on a microfluidic device. *Anal. Chem.* **74**, 3991–4001 (2002).
37. Lee, P. J., Hung, P. J., Shaw, R., Jan, L. & Lee, L. P. Microfluidic application-specific integrated device for monitoring direct cell–cell communication via gap junctions between individual cell pairs. *Appl. Phys. Lett.* **86**, 223902 (2005).
38. Wheeler, A. R. *et al.* Microfluidic device for single-cell analysis. *Anal. Chem.* **75**, 3581–3586 (2003).
39. El-Ali, J., Gaudet, S., Gunther, A., Sorger, P. K. & Jensen, K. F. Cell stimulus and lysis in a microfluidic device with segmented gas-liquid flow. *Anal. Chem.* **77**, 3629–3636 (2005).
40. Song, H., Tice, J. D. & Ismagilov, R. F. A microfluidic system for controlling reaction networks in time. *Angew. Chem. Int. Edn Engl.* **42**, 768–772 (2003).
41. Brehm-Stecher, B. F. & Johnson, E. A. Single-cell microbiology: tools, technologies, and applications. *Microbiol. Mol. Biol. Rev.* **68**, 538–559 (2004).
42. Gascoyne, P. R. C. & Vykuokas, J. V. Dielectrophoresis-based sample handling in general-purpose programmable diagnostic instruments. *Proc. IEEE* **92**, 22–42 (2004).
43. Hu, X., Arnold, W. M. & Zimmermann, U. Alterations in the electrical properties of lymphocyte-T and lymphocyte-B membranes induced by mitogenic stimulation — activation monitored by electro-rotation of single cells. *Biochim. Biophys. Acta* **1021**, 191–200 (1990).
44. Becker, F. F. *et al.* Separation of human breast-cancer cells from blood by differential dielectric affinity. *Proc. Natl Acad. Sci. USA* **92**, 860–864 (1995).
45. Hu, X. Y. *et al.* Marker-specific sorting of rare cells using dielectrophoresis. *Proc. Natl Acad. Sci. USA* **102**, 15757–15761 (2005).
46. Voldman, J., Gray, M. L., Toner, M. & Schmidt, M. A. A microfabrication-based dynamic array cytometer. *Anal. Chem.* **74**, 3984–3990 (2002).
47. Huang, Y. *et al.* Dielectrophoretic cell separation and gene expression profiling on microelectronic chip arrays. *Anal. Chem.* **74**, 3362–3371 (2002).
48. Fu, A. Y., Chou, H. P., Spence, C., Arnold, F. H. & Quake, S. R. An integrated microfabricated cell sorter. *Anal. Chem.* **74**, 2451–2457 (2002).
49. Wang, M. M. *et al.* Microfluidic sorting of mammalian cells by optical force switching. *Nature Biotechnol.* **23**, 83–87 (2005).
50. Chang, W. C., Lee, L. P. & Liepmann, D. Biomimetic technique for adhesion-based collection and separation of cells in a microfluidic channel. *Lab Chip* **5**, 64–73 (2005).
51. Revzin, A., Sekine, K., Sin, A., Tompkins, R. G. & Toner, M. Development of a microfabricated cytometry platform for characterization and sorting of individual leukocytes. *Lab Chip* **5**, 30–37 (2005).
52. Verpoorte, E. Chip vision — optics for microchips. *Lab Chip* **3**, 42N–52N (2003).
53. Wang, Z. *et al.* Measurements of scattered light on a microchip flow cytometer with integrated polymer based optical elements. *Lab Chip* **4**, 372–377 (2004).
54. Li, P. C. H. & Harrison, D. J. Transport, manipulation, and reaction of biological cells on-chip using electrokinetic effects. *Anal. Chem.* **69**, 1564–1568 (1997).
55. Di Carlo, D., Jeong, K. H. & Lee, L. P. Reagentless mechanical cell lysis by nanoscale barbs in microchannels for sample preparation. *Lab Chip* **3**, 287–291 (2003).
56. Lee, S. W. & Tai, Y. C. A micro cell lysis device. *Sens. Actuators A Phys.* **73**, 74–79 (1999).
57. Lu, H., Schmidt, M. A. & Jensen, K. F. A microfluidic electroporation device for cell lysis. *Lab Chip* **5**, 23–29 (2005).
58. McClain, M. A. *et al.* Microfluidic devices for the high-throughput chemical analysis of cells. *Anal. Chem.* **75**, 5646–5655 (2003).
59. Lin, Y. C., Jen, C. M., Huang, M. Y., Wu, C. Y. & Lin, X. Z. Electroporation microchips for continuous gene transfection. *Sens. Actuators B Chem.* **79**, 137–143 (2001).
60. Han, J. & Craighead, H. G. Separation of long DNA molecules in a microfabricated entropic trap array. *Science* **288**, 1026–1029 (2000).
61. Fu, J. P., Mao, P. & Han, J. Y. Nanofilter array chip for fast gel-free biomolecule separation. *Appl. Phys. Lett.* **87**, 263902 (2005).
62. Lu, H., Gaudet, S., Schmidt, M. A. & Jensen, K. F. A microfabricated device for subcellular organelle sorting. *Anal. Chem.* **76**, 5705–5712 (2004).
63. Gottschlich, N., Jacobson, S. C., Culbertson, C. T. & Ramsey, J. M. Two-dimensional electrochromatography/capillary electrophoresis on a microchip. *Anal. Chem.* **73**, 2669–2674 (2001).
64. Li, Y., Buch, J. S., Rosenberger, F., DeVoe, D. L. & Lee, C. S. Integration of isoelectric focusing with parallel sodium dodecyl sulfate gel electrophoresis for multidimensional protein separations in a plastic microfluidic network. *Anal. Chem.* **76**, 742–748 (2004).
65. Wang, Y. C., Stevens, A. L. & Han, J. Y. Million-fold preconcentration of proteins and peptides by nanofluidic filter. *Anal. Chem.* **77**, 4293–4299 (2005).
66. Northrup, M. A., Ching, R. M. & Watson, R. T. in *Proceedings of the 7th International Conference on Solid State Sensors and Actuators* Yokohama, Japan 924–926 (IEEJ, Tokyo, 1993).
67. Burns, M. A. *et al.* An integrated nanoliter DNA analysis device. *Science* **282**, 484–487 (1998).
68. Yin, N. F. *et al.* Microfluidic chip for peptide analysis with an integrated HPLC column, sample enrichment column, and nanoelectrospray tip. *Anal. Chem.* **77**, 527–533 (2005).
69. Sato, K. *et al.* Microchip-based immunoassay system with branching multichannels for simultaneous determination of interferon- γ . *Electrophoresis* **23**, 734–739 (2002).
70. Bernard, A., Michel, B. & Delamar, E. Micromosaic immunoassays. *Anal. Chem.* **73**, 8–12 (2001).
71. Wei, C. W., Cheng, J. Y., Huang, C. T., Yen, M. H. & Young, T. H. Using a microfluidic device for 1 μ l DNA microarray hybridization in 500 s. *Nucleic Acids Res.* **33**, e78 (2005).
72. Gruner, G. Carbon nanotube transistors for biosensing applications. *Anal. Bioanal. Chem.* **384**, 322–335 (2006).
73. Zheng, G. F., Patolsky, F., Cui, Y., Wang, W. U. & Lieber, C. M. Multiplexed electrical detection of cancer markers with nanowire sensor arrays. *Nature Biotechnol.* **23**, 1294–1301 (2005).
74. Burg, T. P. & Manalis, S. R. Suspended microchannel resonators for biomolecular detection. *Appl. Phys. Lett.* **83**, 2698–2700 (2003).
75. Ziegler, C. Cantilever-based biosensors. *Anal. Bioanal. Chem.* **379**, 946–959 (2004).
76. Clayton, J. Go with the microflow. *Nature Methods* **2**, 621–627 (2005).
77. Gu, M. B., Mitchell, R. J. & Kim, B. C. Whole-cell-based biosensors for environmental biomonitoring and application. *Adv. Biochem. Eng. Biotechnol.* **87**, 269–305 (2004).
78. Bousse, L. Whole cell biosensors. *Sens. Actuators B Chem.* **34**, 270–275 (1996).
79. Pancrazio, J. J., Whelan, J. P., Borkholder, D. A., Ma, W. & Stenger, D. A. Development and application of cell-based biosensors. *Ann. Biomed. Eng.* **27**, 697–711 (1999).
80. Gross, G. W., Harsch, A., Rhoades, B. K. & Gopel, W. Odor, drug and toxin analysis with neuronal networks *in vitro*: extracellular array recording of network responses. *Biosens. Bioelectron.* **12**, 373–393 (1997).
81. Pearce, T. M., Wilson, J. A., Oakes, S. G., Chiu, S. Y. & Williams, J. C. Integrated microelectrode array and microfluidics for temperature clamp of sensory neurons in culture. *Lab Chip* **5**, 97–101 (2005).
82. Chang, J. C., Brewer, G. J. & Wheeler, B. C. Modulation of neural network activity by patterning. *Biosens. Bioelectron.* **16**, 527–533 (2001).
83. Morin, F. *et al.* Constraining the connectivity of neuronal networks cultured on microelectrode arrays with microfluidic techniques: a step towards neuron-based functional chips. *Biosens. Bioelectron.* **21**, 1093–1100 (2006).
84. DeBusschere, B. D. & Kovacs, G. T. A. Portable cell-based biosensor system using integrated CMOS cell-cartridges. *Biosens. Bioelectron.* **16**, 543–556 (2001).
85. Wood, C., Williams, C. & Waldron, G. J. Patch clamping by numbers. *Drug Discov. Today* **9**, 434–441 (2004).
86. Fertig, N., Blick, R. H. & Behrends, J. C. Whole cell patch clamp recording performed on a planar glass chip. *Biophys. J.* **82**, 3056–3062 (2002).
87. Schmidt, C., Mayer, M. & Vogel, H. A chip-based biosensor for the functional analysis of single ion channels. *Angew. Chem. Int. Edn Engl.* **39**, 3137–3140 (2000).
88. Seo, J., Ionescu-Zanetti, C., Diamond, J., Lal, R. & Lee, L. P. Integrated multiple patch-clamp array chip via lateral cell trapping junctions. *Appl. Phys. Lett.* **84**, 1973–1975 (2004).
89. Viravaidya, K., Sin, A. & Shuler, M. L. Development of a microscale cell culture array to probe naphthalene toxicity. *Biotechnol. Prog.* **20**, 316–323 (2004).
90. Abhyankar, V. V. & Beebe, D. J. in *Lab-on-Chips for Cellomics* (eds Andersson, H. & van den Berg, A.) 257–272 (Kluwer Academic, Dordrecht, The Netherlands, 2004).
91. Chung, B. G. *et al.* Human neural stem cell growth and differentiation in a gradient-generating microfluidic device. *Lab Chip* **5**, 401–406 (2005).
92. Kim, L., Vahey, M. D., Lee, H. Y. & Voldman, J. Microfluidic arrays for logarithmically perfused embryonic stem cell culture. *Lab Chip* **6**, 394–406 (2006).
93. Pal, R. *et al.* An integrated microfluidic device for influenza and other genetic analyses. *Lab Chip* **5**, 1024–1032 (2005).
94. Lagally, E. T. *et al.* Integrated portable genetic analysis microsystem for pathogen/infectious disease detection. *Anal. Chem.* **76**, 3162–3170 (2004).
95. Blazej, R. G., Kumaresan, P. & Mathies, R. A. Microfabricated bioprocessor for integrated nanoliter-scale Sanger DNA sequencing. *Proc. Natl Acad. Sci. USA* **103**, 7240–7245 (2006).
96. Hong, J. W., Studer, V., Hang, G., Anderson, W. F. & Quake, S. R. A nanoliter-scale nucleic acid processor with parallel architecture. *Nature Biotechnol.* **22**, 435–439 (2004).
97. Wu, H. K., Wheeler, A. & Zare, R. N. Chemical cytometry on a picoliter-scale integrated microfluidic chip. *Proc. Natl Acad. Sci. USA* **101**, 12809–12813 (2004).
98. Gao, J., Yin, X. F. & Fang, Z. L. Integration of single cell injection, cell lysis, separation and detection of intracellular constituents on a microfluidic chip. *Lab Chip* **4**, 47–52 (2004).
99. Geiger, B., Bershadsky, A., Pankov, R. & Yamada, K. M. Transmembrane extracellular matrix–cytoskeleton crosstalk. *Nature Rev. Mol. Cell Biol.* **2**, 793–805 (2001).
100. Jamora, C. & Fuchs, E. Inter cellular adhesion, signalling and the cytoskeleton. *Nature Cell Biol.* **4**, E101–E108 (2002).

Acknowledgements The authors would like to acknowledge funding received from a National Institutes of Health grant.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence should be addressed to K.F.J. (kfjensen@mit.edu).

Microfluidic diagnostic technologies for global public health

Paul Yager¹, Thayne Edwards¹, Elain Fu¹, Kristen Helton¹, Kjell Nelson¹, Milton R. Tam² & Bernhard H. Weigl³

The developing world does not have access to many of the best medical diagnostic technologies; they were designed for air-conditioned laboratories, refrigerated storage of chemicals, a constant supply of calibrators and reagents, stable electrical power, highly trained personnel and rapid transportation of samples. Microfluidic systems allow miniaturization and integration of complex functions, which could move sophisticated diagnostic tools out of the developed-world laboratory. These systems must be inexpensive, but also accurate, reliable, rugged and well suited to the medical and social contexts of the developing world.

Microfluidic systems can be designed to obtain and process measurements from small volumes of complex fluids with efficiency and speed, and without the need for an expert operator; this unique set of capabilities is precisely what is needed to create portable point-of-care (POC) medical diagnostic systems^{1,2}. Fortunately for the microfluidics field, the military has always had a need to practise medicine in challenging and resource-limited environments, and so has long been trying to acquire robust medical technologies that add an absolute minimum to the burden of those people and machines transporting them. It was for this reason that microfluidics research in the United States was given a great boost in the 1990s by funding from the US Defense Advanced Research Projects Agency (DARPA). The technologies developed with DARPA's support (for examples, see www.darpa.mil/MTO/mFlumes) have the characteristics needed for delivering appropriate medical diagnostics to the world's poorest people. Today, the potential of microfluidic technologies to enhance the decentralization of medical testing is becoming accepted as one element in the next stage in the evolution of healthcare. Thanks to an upsurge in interest in (and funding of) healthcare in the developing world, combined with a slow pace of change in the developed world, this new microfluidic diagnostic technology (MDT) may be adopted first for civilian healthcare in the developing world. Some initial steps towards designing appropriate microfluidic diagnostic systems are described here.

Overview of global health issues

In recent decades, substantial progress has been made in public health, but this progress has not been equal in developed and developing countries. The benefits and services that we in the West take for granted are often inadequate or lacking in many developing countries^{3,4}. For example, of the world's population of 6.1 billion people, 3 billion lack basic sanitation, 2 billion do not have access to electricity and more than 1 billion lack basic healthcare services and clean drinking water^{5,6}. For every public health triumph such as the eradication of smallpox, other infectious diseases such as tuberculosis and malaria have re-emerged, accompanied by new diseases such as HIV/AIDS (Fig. 1). More than half the deaths in the poorest countries are the result of infectious diseases (compared with less than 5% in the richest)^{7,8}. Global health, poverty and development are interdependent. Endemic poverty is a significant impediment to improving health⁹. One billion, or half of the world's children, live

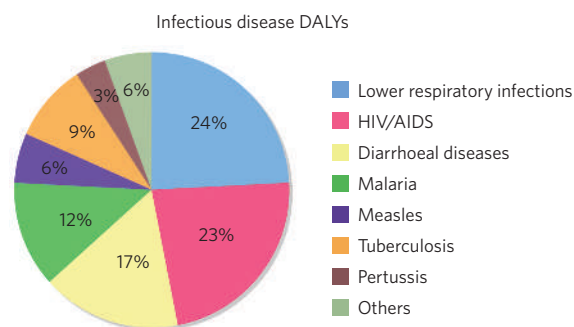


Figure 1 | Disability-adjusted life years (DALYs) for infectious and parasitic diseases. To properly reflect the full impact of a disease, disease burdens can be measured in DALYs by adding the years of life lost by a person's premature death to the time lived with a disability. Infectious and parasitic diseases accounted for almost 30% of all DALYs and 15 million deaths each year worldwide. Shown are the infectious and parasitic diseases responsible for the DALYs in 2005 (figures from the US Centers for Disease Control and Prevention).

in poverty. Health indicators for mothers, newborns and children in many of the poorest countries have remained the same or have even declined in recent years¹⁰.

Until recently, diagnostic tests were not routinely developed primarily for developing-country markets, although two exceptions are the Program for Appropriate Technology in Health (PATH) HealthTech initiative, discussed below, and Helen Lee's diagnostics development programme at the University of Cambridge, UK. Led by the Bill & Melinda Gates Foundation, initial funding has been provided to search for innovative technologies and solutions, including development of new *in vitro* diagnostic tests and test platforms, for the world's neglected diseases¹¹. The US National Institute of Allergy and Infectious Disease is also funding an *in vitro* diagnostics development programme for biodefence to address these issues¹². Three additional initiatives that focus on the development of diagnostic tools are funded by the UK Department for International Development¹³, the Foundation for Innovative New Diagnostics, Geneva, Switzerland (www.finddiagnostics.org) and the Doris Duke Charitable Foundation (www.ddcf.org).

¹Department of Bioengineering, Box 355061, University of Washington, Seattle, Washington 98195, USA. ²7501 21st Avenue NE, Seattle, Washington 98115, USA. ³PATH (Program for Appropriate Technology in Health), 1455 NW Leary Way, Seattle, Washington 98107, USA.

For optimal use in low-resource settings, it is also important for tests to be rapid, simple to use (that is, requiring little in the way of facilities, equipment or training), low-cost or cost-effective, easily interpretable, and stable when transported and stored under extreme conditions. For some infectious agents, tests are needed to better distinguish past from current infections. For others, improved POC methods are needed to return same-day test results so that patients can receive appropriate therapy while they are still at the clinic. Multiplexed tests are especially needed to accurately identify the aetiological agent causing a disease that could have multiple causative agents, such as acute lower respiratory infections, diarrhoeal diseases, acute onset fevers and sexually transmitted infections. Global health programmes are significantly hindered because such tests are currently unavailable. There is also need for individual tests for emerging and re-emerging diseases such as tuberculosis, severe acute respiratory syndrome (SARS) and influenza, as well as vaccine-treatable diseases such as measles, tetanus and polio. For HIV/AIDS, better and more appropriate tests are needed for early diagnosis, case management and treatment monitoring of patients. For tropical diseases, the more urgent needs include those for leishmaniasis, trypanosomiasis and malaria case management and test-of-cure.

Diagnostic tools used in developed-world centralized labs

Microfluidic instrumentation can be and has been applied to several of the four most common centralized laboratory techniques — blood chemistries, immunoassays, nucleic-acid amplification tests and flow cytometry.

Basic blood chemistry panels consisting of 12 to 20 tests are routinely run on automated analysers to monitor a wide range of physiological functions. Analytes include blood enzymes, gases, electrolytes, lipids, thyroid indicators and drugs. Immunoassays in a wide range of formats allow quantification and monitoring of small molecules, large proteins and even whole pathogens. Simple and rapid immunoassays such as lateral flow strip tests (see below) can also be used in clinics or are available over the counter for home use. Blood, plasma, serum, urine, salivary fluids and other exudates are all used as samples. Nucleic-acid amplification tests, such as the polymerase chain reaction (PCR) and nucleic-acid sequence-based amplification, have been developed that can detect very small copy numbers of specific nucleic-acid sequences. Test kits are now commercially available for tuberculosis, HIV and sexually transmitted infections (STIs). Test sensitivity and specificity often exceed that of immunoassays and culture methods¹⁴. 'Real-time' PCR testing is now capable of producing a quantitative result in 20–30 minutes^{15,16}. Flow cytometry is the method of choice for counting cells with specific physical and/or chemical characteristics. Great strides towards miniaturization, including work now published for a decade about microfluidic devices, have been made^{17,18}.

These technologies are well suited to small and large laboratories in the developing world, assuming that the technicians employed to use them have a relatively high level of training. However, in places with lower levels of training, or away from venues with continuous power and, hence, refrigeration, these technologies have been less applicable. Microfluidics may be able to close the gap between what can be done now and what needs to be done at the remotest ends of the healthcare system, where the vast majority of the developing world's population resides.



Figure 2 | Two typical laboratories in mid-level healthcare centres in the developing world. Note the wide range of conditions in which healthcare workers operate. (Images courtesy of B. H. W. and colleagues, PATH, USA.)

End-users of microfluidic diagnostics in the developing world

The centralized laboratory model found in the developed world is not widely applicable in developing nations. Where such centralized laboratories exist at all, most are found in large cities, catering primarily for the affluent few. Healthcare facilities in rural areas commonly have only basic equipment, healthcare workers may have little training, and the resources to maintain complex equipment and handle fragile reagents are often limited. Three tiers of healthcare providers can be identified in developing countries: capital- or major-city-level hospitals; district-level healthcare providers (Fig. 2); and village-level healthcare workers. However, the actual level of care provided at each site can vary considerably.

Typical conditions in a laboratory in a mid-level healthcare centre in a developing country are rather different from those in a developed-world mid-level clinical laboratory. Running water and electricity may or may not be available, but power is at best intermittent with wide fluctuations in voltage. The ambient temperature may range from 10 to more than 40 °C. Dust, wind and contaminating pathogens are very common. Potentially high-risk human samples, containing biosafety level (BSL)-2 and -3 pathogens are routinely handled with few precautions other than, perhaps, gloves. Maintaining and calibrating even moderately complex instruments still presents a challenge. However, it may be possible to perform microscopy, most lateral flow assays, some blood chemistry and cytology, and some enzyme-linked immunosorbent assays (ELISAs) in these labs.

At the periphery of the medical care system — in townships and villages — power, running water and refrigeration are often intermittent or absent. Whatever assays are to be performed must therefore be completely self-contained. Devices must be battery-operated, require no maintenance or calibration, be dust-proof, be easily secured at night, and operate at a wide range of ambient temperatures. Village-level healthcare workers are, for most of the population, the only or main source of healthcare; such health workers can sometimes perform basic strip or dipstick tests for pregnancy, HIV and, in some cases, malaria and STIs. Test results must be presented in a very simple, yes/no way to avoid misinterpretation. To bring a more sophisticated set of diagnostic capabilities to such environments is a challenge of both technical function and development of an appropriate user interface.

Potential benefits of microfluidic diagnostic systems

During the past decade in the developed world, near-patient testing using POC devices has become established for doctors' offices, stat labs and, in rare cases, in the home (for example, glucose monitors). Interest in moving to a more patient-centric POC/home-testing approach is on the rise in the developed world¹⁹. Microfluidic lab-on-a-chip technology may now suit both developing- and developed-world applications. Many of us believe we will soon have microfluidics-based POC and home-care devices that can perform assays at sensitivity, specificity and reproducibility levels similar to those of central laboratory analysers, but yet require little user input other than the insertion of the sample. Such devices could, in the hands of developing-country lay people, perform routine testing, or detect the presence of an infectious agent with epidemic potential such as influenza, an opportunistic infection, or a chronic health condition, and provide guidance to the end-user. That same diagnostic platform, loaded with different reagents, could, in the hands of a village healthcare worker, detect the aetiological agent causing acute fever or diarrhoea, and provide information on what, if any, therapy should be given to the patient. However, the cost of the hypothetical microfluidics-based diagnostic tool must be extremely low if it is to be applicable to the developing world.

Potential benefits of distributing diagnostic devices and systems designed specifically for developing countries include: access to diagnostic tools not previously available, and thus faster/more accurate diagnoses; better epidemiological data that can be used for disease modelling, vaccine introduction and to define the economics of a healthcare system; better utilization of minimally trained healthcare workers; and better use of existing therapeutics.

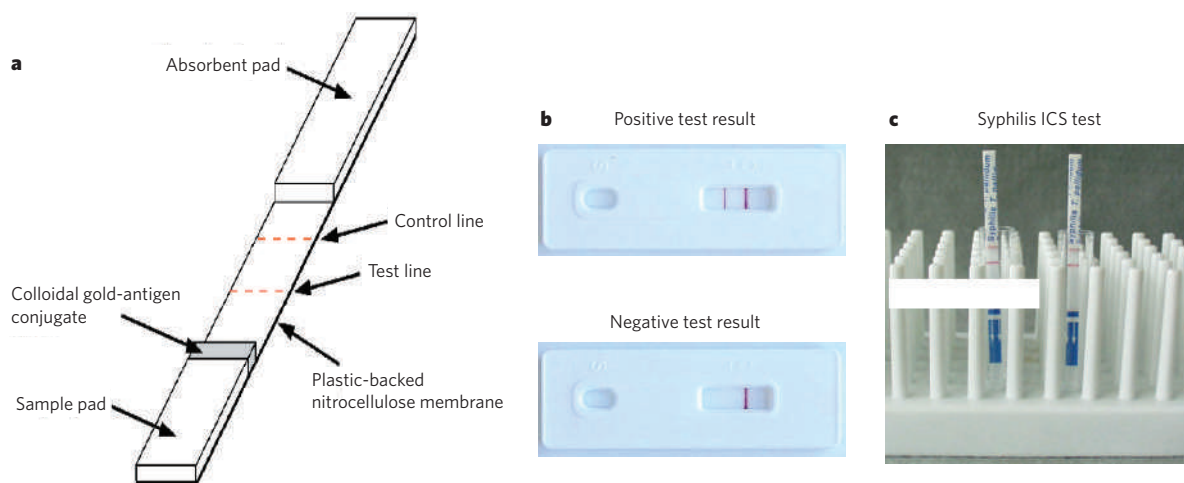


Figure 3 | Rapid immunochromatographic strip (ICS) tests for sexually transmitted infections. **a**, A schematic of the ICS assay format. **b**, **c**, Tests for gonorrhoea (**b**) and syphilis (**c**) are shown.

Choices and challenges in the developing world

There are three approaches to diagnostic technologies that have been used in the developed world: permanent integrated instruments, pure disposables and permanent instruments that use disposable components. All can be and have been adapted for use in the developing world, but they are not equally applicable to specific problems.

Permanent integrated instruments

The permanent integrated instrument is the mainstay of the centralized laboratory in the developed world. These systems are well suited to high-throughput work, but generally require an infrastructure that cannot usually be provided in developing countries. Even if such an instrument could be scaled down (and made affordable enough) to work in resource-limited settings, the requirement that the instrument purge itself of one sample before analysing another (preventing carryover) and that it be frequently calibrated with standards, and rinsed with cleaning solutions, is not in keeping with the setting. This approach is, therefore, unlikely to be successful, particularly in the most impoverished settings, even with microfluidic components.

Disposables

Lateral flow or immunochromatographic strip (ICS) tests have been, for the past decade, one of the very few diagnostic technologies to be successfully used in the developing world, and are, therefore, the technology that microfluidics must complement or supplant. They provide POC diagnosis in areas without access to well-equipped and well-staffed clinical laboratories. They rely on relatively inexpensive, off-the-shelf components and reagents, are amenable to large-scale production methods, and are relatively affordable. They can be formatted for detection of either antigens or antibodies, and are usable with a wide range of specimens. Most are stable at ambient temperatures for more than a year when appropriately packaged and can be shipped without refrigeration. The analytical performance of some POC lateral flow tests are comparable to reference-level laboratory methods³.

ICS tests developed by PATH, with support from the United States Agency for International Development (USAID) can diagnose diphtheria toxin^{20,21–27} and a number of STIs²⁸, including gonorrhoea (*Neisseria gonorrhoeae*) (Fig. 3), syphilis (*Treponema pallidum*)²⁹, chancroid (*Haemophilus ducreyi*)²⁰ and chlamydia (*Chlamydia trachomatis*)^{28,30–32}. The early diagnosis offered by some strip tests, together with the ability to provide appropriate therapy to patients before they leave the clinic, has proved useful to control the spread of the diseases³³. In addition to STI assays, PATH has developed, co-developed, or introduced assays for retinol-binding protein (as a marker for vitamin A deficiency)³⁴, *P. falciparum* malaria^{30–32}, HIV^{25,29}, hepatitis B²⁴, the hormone human

chorionic gonadotropin (indicating pregnancy), faecal leukocytes and proteinuria^{20–23,26,28}.

Health workers can quickly learn to perform such ICS-based tests and require infrequent retraining. However, some such tests are still not sufficiently sensitive or specific for accurate POC use. For example, PATH's malaria test, which can only detect *P. falciparum*³¹, must be used with care in hyperendemic areas, where many individuals may have low titres of circulating *P. falciparum* antigens, leading to an excess of false positive results. Parasite antigens may persist in circulation even after successful therapy, and could produce confusing results with *P. falciparum* tests when assessing test-of-cure or drug resistance. The visual readout of the strip is usually limited to a yes/no answer; this is not adequate when the level of an analyte is important. Sophisticated sample preconditioning (which is needed for many assays) is also impossible in many areas.

Disposables with a reader

The compromise that allows high performance with low per-test cost, but with added complexity, is to use a 'reader' with single-use disposables. The sample and waste are retained within the disposable, so the reader does not need to be cleaned between samples. Calibrants can also be stored on the disposable. At the University of Washington, we are developing a system for monitoring analytes in saliva using surface plasmon resonance (SPR) imaging^{35–37}. The aim of the salivary diagnostics system is to measure small-molecule analytes such as hormones and drugs in whole saliva. The use of saliva is a suitably challenging problem, and one that would translate well to use in the developing world. The 'disposables with a reader' model seems to us to be the most promising approach as a powerful and versatile format that can meet the demands of the developing-world setting. Below, we consider in more detail the challenges to overcome for this diagnostic platform.

The first challenge is to process complex biological samples without the sophisticated sample preconditioning capabilities (human or mechanical) available in centralized labs. Centrifugation, which might be available in a mid-level laboratory to provide a purified plasma sample from blood, will not be available in a remote setting. There are uniquely microfluidic solutions to these issues, such as the H-filter^{38–41}, which was conceived as an alternative to a conventional porous barrier filter (Fig. 4). The H-filter can be used to limit the size of analytes that proceed downstream to a detection module from non-blood samples such as saliva, at the cost of requiring controlled flow to three of the four ports of the device. However, for single-use disposables, a conventional filter may be adequate for removing cellular components from blood prior to analysis. In the case of the salivary diagnostics system, it was found that saliva contained molecules that rapidly bound to the SPR detection surface, adding unacceptable background signal, so

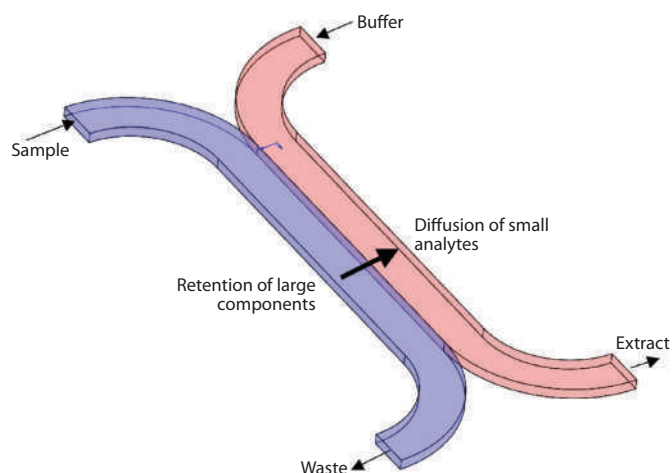


Figure 4 | Schematic of an H-filter. This is a microfluidic device that allows convenient extraction of small molecules from complex fluids into simpler buffer streams. Initial attempts to use the H-filter alone to extract small-molecule analytes from whole saliva failed because of the high viscosity and non-newtonian behaviour of the solution. It was therefore necessary that the concentration of mucins be reduced before the saliva entered the H-filter (manuscript in preparation, K.H., B. Finlayson and P.Y.). Note also that in the disposable card in Fig. 5, the H-filter channels were configured to allow greater contact between the two flowing streams.

some form of sample preconditioning was necessary. Electrophoresis, dielectrophoresis and isoelectric focusing are three other techniques that can also be used, and have been used extensively in lab-on-a-chip systems^{42–51}. The electrokinetic methods can add cost if they require electrodes to be present in the disposable itself.

One of the great strengths of microfluidics is the ability to integrate the steps of a complex chemical process into a monolithic disposable — an assay performed in such a way can be more precise, more accurate and more reproducible than the same assay performed by hand. An example of an integrated assay is shown in Fig. 5, in which many of the steps of analysis of a whole saliva sample (subsequent to mechanical filtration) are integrated into a disposable diagnostic card. In this card, the sample is isolated and the waste is retained, and sample contact with the electrical and optical components of the permanent instrument is avoided. Note that not all of the challenges of integration have been accomplished in the current design; for example, external pumps and off-chip sample preprocessing are required. However, two other projects that are underway (and described briefly below) will produce disposables that fully integrate all fluidic functions.

The selection of a material for use in a disposable must balance the inherent cost of the material with processing costs. Many polymers are inherently inexpensive, but the cost of processing to make them compatible with device function can add substantially to the total cost of the disposable. For example, microfluidic materials must be selected to minimize surface adsorption of analytes and to allow those analytes to move to the detection zones in the disposables. A material can be inherently nonfouling, but more often must undergo a coating process to be made so. An inherently low-cost material may be more difficult to chemically modify and would increase the complexity of manufacturing and the final cost of the disposable.

Often, selection of a material will be dictated by application requirements. For example, in applications that require fluorescence-based detection of small-molecule analytes, low fluorescence (and low light scattering) is a selection criterion for acceptable device materials in the optical path⁵². One material, poly(dimethylsiloxane) (PDMS), is excellent for fabrication of low-fluorescence devices, but is permeable to many small and hydrophobic molecules, and is not readily or economically formed in high-throughput production. Thermoplastics such as polystyrene are ideal for high-level production because they can

be injection molded, although not all have optimal optical properties. Whereas lamination of laser-cut polymer sheets is extremely cost-effective for rapid prototyping⁵³, it is not nearly as economical as injection moulding for production of large numbers of devices. Although mixing different materials is complex, it may often be necessary. Note that in the laminated disposable example (Fig. 5a), PDMS was used to form a herringbone mixer⁵⁴ with small features, largely because the CO₂ laser (used to form the Mylar layers) had too low a spot resolution. Also, the gold SPR imaging surface was formed on a glass microscope slide because it was less expensive to manufacture them this way than to purchase small lots of gold films of a precise thickness on polymers. The fragility of glass makes it unsuitable for use in a commercializable disposable for POC diagnostics. Real devices will always have to combine multiple materials.

Because one of the fundamental requirements for a microfluidic diagnostic disposable is extremely low cost (ideally pennies), the disposables must be designed to have few parts and be inexpensive to manufacture, ideally using injection moulding whenever possible. These criteria rule

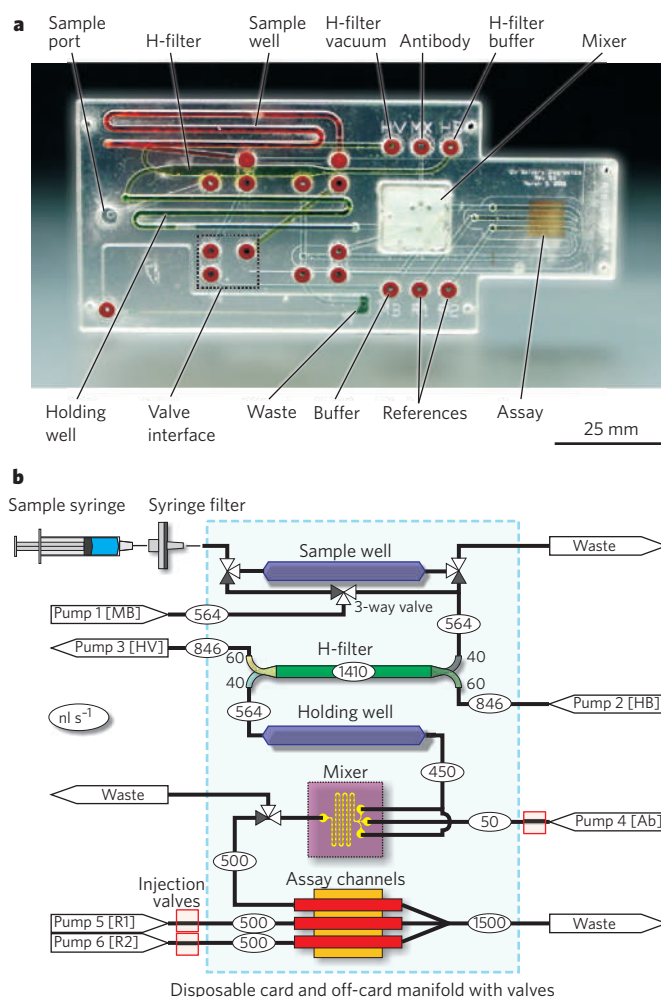


Figure 5 | Example of an integrated disposable diagnostic card. **a**, Image of a card. The red O-rings are for interfacing with off-card components (for example, valves and pumps) that will eventually be incorporated onto the card itself. **b**, Schematic of the card. The card accepts filtered saliva (see Table 2) from the syringe and contains an H-filter for further sample conditioning, a herringbone mixer for mixing antibodies with the sample, and channels with gold-coated surfaces for detection of analyte in the sample using an SPR imaging-based immunoassay (manuscript in preparation, T.E. *et al.*). Numbers in ovals are flow rates in nl s⁻¹. Numbers at the H-filter's four ports are the percentages of flow entering (right) and exiting (left) the device. MB, mixing buffer; HB, H-filter buffer; HV, H-filter vacuum; R1 and R2 are reference solutions, typically a positive and negative control.

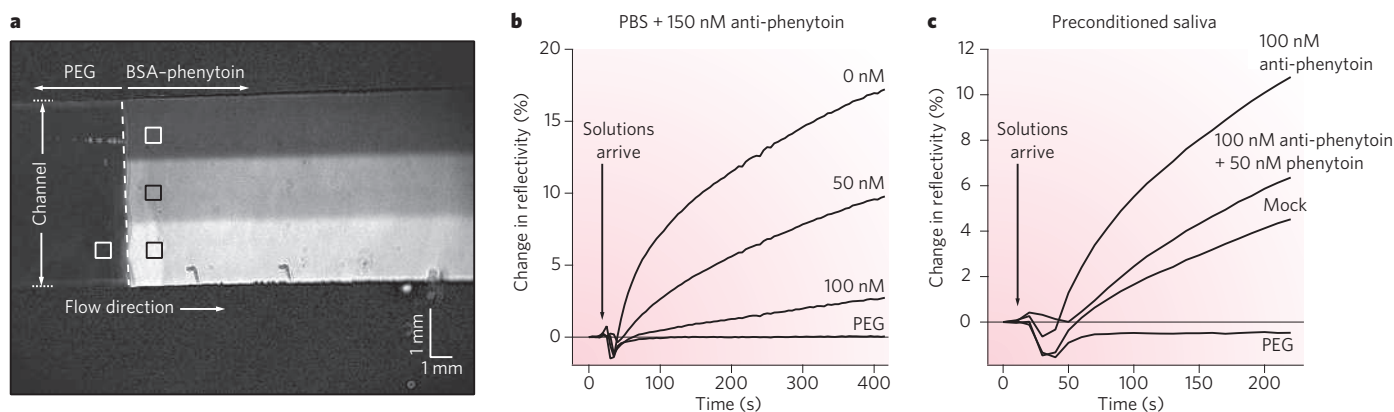


Figure 6 | Quantification of a competitive immunoassay for phenytoin using surface plasmon resonance (SPR). **a**, SPR difference image used to quantify a competitive immunoassay for phenytoin (manuscript in preparation, K.N. *et al.*). The detection zone is a gold surface precoated with bovine serum albumin (BSA) covalently modified by phenytoin. The three streams (750 nl s⁻¹) contained 150 nM anti-phenytoin antibody mixed with 0 nM, 50 nM or 100 nM soluble phenytoin (bottom to top, respectively). The contrast in the image has been adjusted to highlight the differences between the nonfouling polyethylene glycol (PEG) upstream and the three sample regions. **b**, **c**, Competition assay detection of low-end therapeutic levels of phenytoin in a model system (**b**) and in preconditioned saliva (**c**). **b**, SPR reflectivity over time of anti-phenytoin antibody in phosphate buffer binding to a BSA-phenytoin-treated surface. The plot shows that the rate of antibody binding negatively correlates with the amount of competitor (phenytoin) added to the solution. **c**, SPR reflectivity of variously treated preconditioned saliva samples over time. In this case, whole human saliva was preconditioned using a mechanical filter and the H-filter off-card. Plots **b** and **c** show that the PEG region effectively resisted fouling to either antibody or components present in preconditioned saliva.

out inherently expensive materials, complex manufacturing methods and expensive reagents in anything but the smallest quantities. It is also advisable, on cost grounds alone, to assemble and manufacture as many parts of the disposable as possible in the developing world.

The disposable/reader model results in the creation of biohazardous medical waste of little or no value for recycling. The best method for disposing of such waste is incineration, but it is also possible to render it non-infectious by treatment with a disinfectant such as bleach, which could even be incorporated into the disposable in a blister pack. For use in remote areas, thought should therefore be given to producing an efficient, safe and verifiable method for incinerating disposables using local facilities.

The microfluidic diagnostic end-users will need a supply of disposables. The disposables will need to be robust, as they will be transported at ambient temperatures. Because the disposables will be hand-carried, their weight must be minimal. Large, centralized instruments frequently require large volumes of pre-made buffers and reagent solutions to operate. Fluid reservoirs contribute substantially to the bulk of the instrumentation (and the revenue stream of the manufacturers). Because water is heavy, there is a clear incentive to ship the disposables dry, and use local sources of purified water.

Although lateral flow strips and related technologies do not require power to operate, instruments currently used in centralized laboratories are designed with no consideration of their electrical requirements. For a field-portable instrument, the assumption is that electrical power will either be nonexistent, or limited to automobile generators, photocells, hand-generators, or other low-capacity power sources. As a consequence, MDT systems must be designed from the ground up to use a minimum amount of power, and if the disposables store power, it must be in clever ways — a challenging task for any but the simplest (and least quantitative) assays. Such high-precision assays — for example, those requiring thermal cycling (such as PCR and related nucleic-acid amplification technologies) — require a lot of power in their current embodiments, so there is room for creative engineering to develop inherently low-power methods of changing sample temperature. What power sources may be used (other than external ones) must be compact and lightweight. Great advances made during the last few years in small, rechargeable batteries will help, but there is clearly scope for improvement in this area too.

A successful microfluidic diagnostic technology will be self-calibrating, carrying all necessary reagents to run positive and negative controls along with it. This is particularly necessary for a POC instrument, which

will be especially sensitive to variations in ambient temperature because of its small size. Ideally, these controls will be present on the disposable itself, so that every measurement will be made at the same time as the calibration runs. This will put a heavy burden on the preservation of the reagents in the disposable. In the salivary diagnostics project, we have developed microfluidic methods for running calibrations alongside the critical measurements by patterning reagents and using multiple low-Reynolds-number fluid streams to keep detection areas discrete. Such methods allow measurement of binding rates of analytes to surfaces in samples such as saliva in just a few minutes (Fig. 6).

The disposables themselves will be shipped and stored at ambient temperatures, which, in the developing world, can range from tropical to arctic. For vaccines, which are generally very heat-labile, the 'cold chain' has been extended to the periphery of the healthcare system. However, it is costly to maintain. The disposables for diagnostics must contain biological molecules (and perhaps even entire pathogens) in addition to a range of organic chemicals as reagents. If they were stored in buffer at ambient temperature they would very rapidly be degraded. A preferred alternative, successfully used in lateral flow assays and recently applied to more readily identifiable 'microfluidic' systems, is to desiccate the biomolecules, replacing the normal waters of hydration by sugars^{55,56}. This, derived in part from the drug-delivery field, would allow the disposable to contain only dry reagents, thus reducing weight, extending product stability and allowing transportation and storage of the MDTs at ambient temperature.

A critical aim of microfluidic diagnostics for the developing world is to make all steps of their use simple and as culturally independent as possible. This will be a challenge given the diversity of language, training and cultural backgrounds of the end-users of the technologies. For this reason, extensive work with the end-users is an absolute requirement before designing the user interface. It is not possible to assume that the end-user will be the same as the users of centralized laboratory equipment.

The future of MDTs in the developing world

Enteric infections are the second leading cause of morbidity and mortality worldwide, accounting for an estimated 3.1 million deaths annually, mostly in the developing world. Outbreaks can be controlled with rapid diagnosis and appropriate treatment, which can also reduce the severity of disease. But the developed-world standards for diagnosis of infectious diarrhoea — culture, enzyme immunoassay and PCR — are impractical, expensive and too slow for developing-world users. Identification of target pathogens, even in the best developed-world laboratory, often takes 2–4 days⁵⁷.

Table 1 | Factors that will affect the use of a microfluidic diagnostic technology (MDT) and the attributes of a successful MDT in the developing world

Key factors that will affect the introduction, acceptance and sustained use of an MDT	Key attributes of an MDT
Cost of the technology	Low cost
Degree of accuracy	High degree of accuracy (for example, low rate of false positives)
Quality control	Reproducible chip performance (using on-chip calibrators)
Level of training of users	User interface that requires little training
Length of time to obtain test result	Short time to test result
Performance in a variety of settings	Stable ambient temperature storage and low power consumption
Performance under variable operating conditions (such as temperature or humidity) over time	Reproducible operation in variable environments, and ruggedness
Local education on health issues	A high perceived need for the test
Availability of successful therapies	Potential for significant health improvements

Work has begun on a 'disposable with reader' lab-on-a-chip platform for identification of the pathogens — *Shigella dysenteriae* type 1, Shiga toxin-producing *Escherichia coli* (O157:H7), *Campylobacter jejuni* and *Salmonella* — that commonly cause acute enteric disease with similar symptoms. The diagnostic assay comprises the disposable single-use microfluidic card, containing dry, heat-stable agents and a permanent hand-held instrument to operate the microfluidic circuits and control the card's temperature. The user will insert a swab containing a stool sample into the card and place the card in the instrument. The card will include four microfluidic subcircuits: organism capture and lysis from raw stool; nucleic-acid capture; multiplexed nucleic-acid amplification; and visual detection of amplified PCR products. A combination of capillary action and positive displacement pumping will draw the sample via microfluidic channels through the integrated subcircuits on the disposable. A positive control (*E. coli*, present in any stool sample) will be included to demonstrate proper sample processing, and to validate negative results. The complete sequence will take less than 30 min. Early tests show that sensitivity and specificity are comparable to the results achieved with conventional microbiological and PCR assays⁵⁸, and the cost is expected to be between US\$1 and \$5 per disposable.

The Gates Foundation's Grand Challenges in Global Health initiative is supporting the development of prototypes of a disposable/ hand-held reader system, which will include the reagents for both immunoassays and nucleic-acid amplification tests, requiring only a couple of drops of blood from the patient for a readout. Initial focus is on the development of tests for simultaneous detection of infectious agents that cause diseases associated with rapid onset fever, including malaria, typhoid, dengue, rickettsial diseases, measles and influenza. Because of the ability to perform both immunoassays and nucleic-acid amplified tests, this platform could ultimately process a wide range of disease panels.

The current funding climate for the development of diagnostics for developing countries is good, thanks to vastly increased interest in global health, stoked by initiatives from private foundations and institutions. Additionally, significant funding of pathogen diagnostics development is available through United States government sources¹², although primarily driven by biodefense concerns, as many of the organisms that could be weaponized also cause disease in developing countries (www3.niaid.nih.gov/biodefense/bandc_priority.htm). Donors and international agencies support the need for new and improved diagnostic tools for priority diseases; advocacy and technology development groups have been formed, such as the STD Diagnostics Initiative⁵⁹ and the Tuberculosis Diagnostics Initiative⁶⁰, with secretariats within the World Health Organisation/Special Programme for Research and Training

in Tropical Diseases (WHO/TDR). These are good beginnings, but because the marketplace for medical diagnostics is fragmented in the developing world, no single strategy will ensure distribution of new technologies as they are developed.

Although we feel strongly that the development of appropriate diagnostic technologies is an important factor in the goal of improving global public health, significant improvements in the health of the developing world will only be achieved if there is tight coordination between the diagnostics developers and the communities involved in local education, drug discovery and drug distribution. Long-term, successful use of MDTs in the developing world will require the sustained efforts of these communities.

There are a number of key factors that will affect the introduction, acceptability and sustainability of these technologies (summarized in Table 1). One of the greater challenges in deploying microfluidic diagnostic systems in the developing world will be bringing the cost down close to the cost of the most inexpensive of current tests, namely the lateral flow immunoassays (it should be borne in mind that the actual cost of using these immunoassays does include that of misdiagnosing patients). Thanks to increased interest on the part of the global health community, we expect to see the introduction of microfluidic diagnostic devices specifically designed for the developing world within the next 5 years.

- Dupuy, A., Lehmann, S. & Cristol, J. Protein biochip systems for the clinical laboratory. *Clin. Chem. Lab. Med.* **43**, 1291-1302 (2005).
- Toner, M. & Irimia, D. Blood-on-a-chip. *Annu. Rev. Biomed. Eng.* **7**, 77-103 (2005).
- von Lode, P. Point-of-care immunotesting: approaching the analytical performance of central laboratory methods. *Clin. Biochem.* **38**, 591-606 (2005).
- World Bank World Development Report 1993: Investing in Health (Oxford Univ. Press, New York, 1993).
- Black, R. E., Morris, S. S. & Bryce, J. Where and why are 10 million children dying every year? *Lancet* **361**, 2226-2234 (2003).
- World Health Organization World Health Report 2002 — Reducing Risks, Promoting Healthy Life [online] <www.who.int/whr/2002/en/index.html> (2002).
- Morens, D. M., Folkers, G. K. & Fauci, A. S. The challenge of emerging and re-emerging infectious diseases. *Nature* **430**, 242-249 (2004).
- Fauci, A. S. Infectious diseases: considerations for the 21st century. *Clin. Infect. Dis.* **32**, 675-685 (2001).
- National Research Council Committee on Science and Technology in Foreign Assistance *The Fundamental Role of Science and Technology in International Development: An Imperative for the US Agency for International Development* (Washington DC, 2006).
- Becker, G. S., Philipson, T. J. & Soares, R. R. The quantity and quality of life and the evolution of world inequality. *Am. Econ. Rev.* **95**, 277-291 (2005).
- Murray, C. J. L. & Lopez, A. D. *The Global Burden of Disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020* (Harvard Univ. Press, Cambridge, Massachusetts, 1996).
- Conroy, J. Developing biodefense IVDs still a priority. *IVD Technol.* **12**, 37-46 (2006).
- UK Department for International Development *The Challenge of TB and Malaria Control* [online] <<http://www.dfid.gov.uk/pubs/files/tb-malaria-control.pdf>> (2005).
- Jalal, H., Stephen, H., Al-Suwaine, A., Sonnex, C. & Carne, C. The superiority of polymerase chain reaction over an amplified enzyme immunoassay for the detection of genital chlamydial infections. *Sex. Transm. Infect.* **82**, 37-40 (2006).
- Boel, C. H., van Herk, C. M., Berretty, P. J., Onland, G. H. & van den Brule, A. J. Evaluation of conventional and real-time PCR assays using two targets for confirmation of results of the COBAS AMPLICOR *Chlamydia trachomatis*/Neisseria gonorrhoeae test for detection of Neisseria gonorrhoeae in clinical samples. *J. Clin. Microbiol.* **43**, 2231-2235 (2005).
- Eickhoff, M. et al. Ultra-rapid detection of *Chlamydia trachomatis* by real-time PCR in the LightCycler using SYBR green technology or 5'-nuclease probes. *Clin. Lab.* **49**, 217-225 (2003).
- Dittrich, P. S. & Manz, A. Single-molecule fluorescence detection in microfluidic channels—the Holy Grail in μ TAS? *Anal. Bioanal. Chem.* **382**, 1771-1782 (2005).

Table 2 | Effectiveness of preconditioning methods for salivary diagnostics

Sample	Whole saliva	Filtrate	Extract
Mucin/glycoprotein	100%	27%	2%
Total protein	100%	55%	9%
Analyte (Cortisol)	100%	92%	27%

Mucin, total protein and analyte content for whole human saliva, the filtrate (after the use of the depth filter) and the extract (after extraction with the H-filter).

18. Rodriguez, W. R. *et al.* A microchip CD4 counting method for HIV monitoring in resource-poor settings. *PLoS Med.* **2**, e182 (2005).
19. Price, C. P. Regular review: point of care testing. *Brit. Med. J.* **332**, 1285–1288 (2001).
20. Patterson, K. *et al.* Development of a rapid immunodiagnostic test for *Haemophilus ducreyi*. *J. Clin. Microbiol.* **40**, 3694–3702 (2002).
21. Engler, K. H. *et al.* Immunochromatographic strip test for rapid detection of diphtheria toxin: description and multicenter evaluation in areas of low and high prevalence of diphtheria. *J. Clin. Microbiol.* **40**, 80–83 (2002).
22. Zarakolu, P., Buchanan, I., Tam, M., Smith, K. & Hook, E. W. Preliminary evaluation of an immunochromatographic strip test for specific *Treponema pallidum* antibodies. *J. Clin. Microbiol.* **40**, 3064–3065 (2002).
23. Talbot, E. A. *et al.* Tuberculosis serodiagnosis in a predominantly HIV-infected population of hospitalized patients with cough, Botswana, 2002. *Clin. Infect. Dis.* **39**, 1–7 (2004).
24. Raj, A. A., Subramaniam, T., Raghuraman, S. & Abraham, P. Evaluation of an indigenously manufactured rapid immunochromatographic test for detection of HBsAg. *Indian J. Pathol. Microbiol.* **44**, 413–414 (2001).
25. Kaur, H., Dhanao, J. & Oberoi, A. Evaluation of rapid kits for detection of HIV, HBsAg and HCV infections. *Indian J. Med. Sci.* **54**, 432–434 (2000).
26. Parkes, R., Lo, T., Wong, Q., Isaac-Renton, J. L. & Byrne, S. K. Comparison of a nested polymerase chain reaction–restriction fragment length polymorphism method, the PATH antigen detection method, and microscopy for the detection and identification of malaria parasites. *Can. J. Microbiol.* **47**, 903–907 (2001).
27. Cates, W. Preserving fertility: an underappreciated aspect of sexual health. *Network* **23**, no. 2 (2003).
28. Tam, M. R. in *Sexually Transmitted Diseases* (eds Holmes, K. K. *et al.*) 1409–1420 (McGraw-Hill, New York, 1999).
29. Cayemittes, M., Hankins, C. & Tam, M. R. An AIDS test that travels well. *IDRC Rep.* **21**, 27–28 (1993).
30. Loutfy, M. R., Assmar, M., Hay Burgess, D. C. & Kain, K. C. Effects of viral hemorrhagic fever inactivation methods on the performance of rapid diagnostic tests for *Plasmodium falciparum*. *J. Infect. Dis.* **178**, 1852–1855 (1998).
31. Mills, C. D., Hay Burgess, D. C., Taylor, H. J. & Kain, K. C. Evaluation of a rapid and inexpensive dipstick assay for the diagnosis of *Plasmodium falciparum* malaria. *Bull. World Health Organ.* **77**, 553–559 (1999).
32. Labbe, A. C. *et al.* The performance and utility of rapid diagnostic assays for *Plasmodium falciparum* malaria in a field setting in the Lao People's Democratic Republic. *Ann. Trop. Med. Parasitol.* **95**, 671–677 (2001).
33. Vickerman, P., Peeling, R. W., Watts, C. & Mabey, D. Detection of gonococcal infection: pros and cons of a rapid test. *Mol. Diagn.* **9**, 175–179 (2005).
34. Hix, J. *et al.* Development of a rapid enzyme immunoassay for the detection of retinol-binding protein. *Am. J. Clin. Nutr.* **79**, 93–98 (2004).
35. Chinowsky, T. *et al.* Compact surface plasmon resonance imaging system for saliva-based medical diagnostics. *Biosens. Bioelectron.* (submitted).
36. Fu, E., Chinowsky, T., Foley, J., Weinstein, J. & Yager, P. Characterization of a wavelength-tunable surface plasmon resonance microscope. *Rev. Sci. Instrum.* **75**, 2300–2304 (2004).
37. Fu, E., Foley, J. & Yager, P. Wavelength-tunable surface plasmon resonance microscope. *Rev. Sci. Instrum.* **74**, 3182–3184 (2003).
38. Yager, P., Holl, M. R., Weigl, B. H. & Brody, J. P. (Univ. Washington, 1998).
39. Brody, J. P. & Yager, P. Diffusion-based extraction in a microfabricated device. *Sens. Actuators A Phys.* **A58**, 13–18 (1997).
40. Brody, J. P., Yager, P., Goldstein, R. E. & Austin, R. H. Biotechnology at low Reynolds numbers. *Biophys. J.* **71**, 3430–3441 (1996).
41. Brody, J. P., Osborn, T. D., Forster, F. K. & Yager, P. A planar microfabricated fluid filter. *Sens. Actuators A Phys.* 704–708 (1996).
42. Gascoyne, P. R. C. & Vykhukal, J. Particle separation by dielectrophoresis. *Electrophoresis* **23**, 1973–1983 (2002).
43. Gascoyne, P. *et al.* Microsample preparation by dielectrophoresis: isolation of malaria. *Lab Chip* **2**, 70–75 (2002).
44. Wang, X. B. *et al.* Cell separation by dielectrophoretic field-flow-fractionation. *Analyt. Chem.* **72**, 832–839 (2000).
45. Müller, T., Schnelle, T., Gradl, G., Shirley, S. G. & Fuhr, G. Microdevice for cell and particle separation using dielectrophoretic field-flow fractionation. *J. Liq. Chromatogr. Relat. Technol.* **23**, 47–59 (2000).
46. Hughes, M. AC electrokinetics applications for nanotechnology. *Nanotechnology* **11**, 124–132 (2000).
47. Rousselet, J., Markx, G. H. & Pethig, R. Separation of erythrocytes and latex beads by dielectrophoretic levitation and hyperlayer field-flow fractionation. *Colloids Surf. A* **140**, 209–216 (1998).
48. Cabrera, C. R., Macounová, K., Holl, M. R. & Yager, P. in *IEEE 1st Annual Conference on Microtechnology in Medicine and Biology* (ed. Beebe) (IEEE Press, Lyon, 2000).
49. Cabrera, C. R. & Yager, P. Continuous concentration of bacteria in a microfluidic flow cell using electrokinetic techniques. *Electrophoresis* **22**, 355–362 (2001).
50. Macounová, K., Cabrera, C. R. & Yager, P. Concentration and separation of proteins in microfluidic channels on the basis of transverse IEF. *Analyt. Chem.* **73**, 1627–1633 (2001).
51. Cheng, S. B. *et al.* Development of a multichannel microfluidic analysis system employing affinity capillary electrophoresis for immunoassay. *Analyt. Chem.* **73**, 1472–1479 (2001).
52. Hawkins, K. R. & Yager, P. Nonlinear decrease of background fluorescence in polymer thin-films — a survey of materials and how they can complicate fluorescence detection in μ TAS. *Lab Chip* **3**, 248–252 (2003).
53. Yager, P. *et al.* in *Micro Total Analysis Systems '98* (eds Harrison, D. J. & van den Berg, A.) 207–212 (Kluwer Academic, Dordrecht, Banff, 1998).
54. Stroock, A. *et al.* Chaotic mixer for microchannels. *Science* **295**, 647–651 (2002).
55. Garcia, E., Kusmanto, F., Finlayson, B. & Yager, P. in *Micro Total Analysis Systems 2003* (eds Northrup, M. A., Jensen, K. F. & Harrison, D. J.) 551–554 (Mesa Monographs, Enschede, 2003).
56. Garcia, E., Kirkham, J. R., Hatch, A. V., Hawkins, K. R. & Yager, P. Controlled microfluidic reconstitution of functional protein from an anhydrous storage depot. *Lab Chip* **4**, 78–82 (2004).
57. Guerrant, R. L. *et al.* Practice guidelines for the management of infectious diarrhea. *Clin. Infect. Dis.* **32**, 331–351 (2001).
58. Weigl, B. *et al.* Fully integrated multiplexed lab-on-a-card assay for enteric pathogens. *Proc. SPIE* **6112**, 1–11 (2006).
59. Tam, M. R. & Holmes, K. K. The STD diagnostic network: priorities for diagnosis of STDs in resource-poor settings. *Ups. J. Med. Sci. Suppl.* **50**, 22–23 (1991).
60. Perkins, M. D. & Kritski, A. L. Diagnostic testing in the control of tuberculosis. *Bull. World Health Organ.* **80**, 512–513 (2002).

Acknowledgements We thank our colleagues at PATH, Micronics, Nanogen, Invetech, the University of Washington, Seattle, and Washington University in St Louis, Missouri, for their many contributions to ongoing and past projects, and our funding sources, currently including the National Institutes of Health (the National Institute of Allergy and Infectious Diseases, and the National Institute of Dental and Craniofacial Research) and the Bill & Melinda Gates Foundation. Thanks to colleagues at PATH for the photographs of laboratories in the developing world.

Author Information Reprints and permission information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence should be addressed to P.Y. (yagerp@u.washington.edu).

Snapshots of tRNA sulphuration via an adenylated intermediate

Tomoyuki Numata¹, Yoshiho Ikeuchi², Shuya Fukai¹, Tsutomu Suzuki² & Osamu Nureki^{1,3,4}

Uridine at the first anticodon position (U34) of glutamate, lysine and glutamine transfer RNAs is universally modified by thiouridylase into 2-thiouridine (s^2 U34), which is crucial for precise translation by restricting codon-anticodon wobble during protein synthesis on the ribosome. However, it remains unclear how the enzyme incorporates reactive sulphur into the correct position of the uridine base. Here we present the crystal structures of the MnmA thiouridylase-tRNA complex in three discrete forms, which provide snapshots of the sequential chemical reactions during RNA sulphuration. On enzyme activation, an α -helix overhanging the active site is restructured into an idiosyncratic β -hairpin-containing loop, which packs the flipped-out U34 deeply into the catalytic pocket and triggers the activation of the catalytic cysteine residues. The adenylated RNA intermediate is trapped. Thus, the active closed-conformation of the complex ensures accurate sulphur incorporation into the activated uridine carbon by forming a catalytic chamber to prevent solvent from accessing the catalytic site. The structures of the complex with glutamate tRNA further reveal how MnmA specifically recognizes its three different tRNA substrates. These findings provide the structural basis for a general mechanism whereby an enzyme incorporates a reactive atom at a precise position in a biological molecule.

Post-transcriptional chemical modification of cellular RNAs is widespread and serves diverse functions. Thiouridines appear in tRNAs as 2-thiouridine (s^2 U34) in the anticodon wobble position and 4-thiouridine (s^4 U8) at the junction of the acceptor and D stems. The s^2 U34 modification is phylogenetically conserved in tRNA^{Glu}, tRNA^{Lys} and tRNA^{Gln} from all organisms^{1,2}. Glutamate, lysine and glutamine are encoded by two degenerate codons ending in purine (NNR) in the two-codon boxes that specify two amino acids by difference in the third codon bases in the genetic code. In the messenger RNA decoding site of the 30S ribosome, 2-thio allows U34 to adopt the C3'-endo puckering conformation of the ribose, facilitating base pairing only with purines³ and preventing incorrect decoding of the NNY codon, thereby contributing to fidelity⁴⁻⁶. Furthermore, s^2 U34 serves as a determinant for tRNA recognition by cognate aminoacyl-tRNA synthetases⁷. The 2-thio modification of tRNA^{Lys} is also crucial for assembly of the viral genomic RNA of human immunodeficiency virus type 1 (HIV-1) and tRNA^{Lys}, which is used by the viral reverse transcriptase as a primer^{8,9}.

Escherichia coli MnmA¹⁰ thiouridylase (M_r 41,000) specifically synthesizes the s^2 U34 modifications of tRNA^{Glu}, tRNA^{Lys} and tRNA^{Gln} (ref. 11). The first step of the reaction is reductive elimination of sulphur from L-cysteine by IscS cysteine desulphurase to form an enzyme-bound cysteine-persulphide intermediate. Then, five essential gene products—TusA, TusB, TusC, TusD and TusE—mediate a sulphur relay that delivers the terminal sulphur of persulphide from IscS to MnmA¹². The reactivity of sulphur indicates that MnmA possesses a sophisticated mechanism to ensure specific modification of U34 at the correct position. The catalytic mechanism of RNA sulphuration by MnmA, as well as its modes of tRNA recognition and substrate selection, is unknown.

Here we report the crystal structures of three different forms of the *E. coli* MnmA-tRNA^{Glu} complex. The three co-crystal structures provide snapshots of the sequential chemical reactions before and

after adenylation of U34: the initial tRNA-binding (form I), the pre-reaction (form II), and the adenylated intermediate states (form III) (Supplementary Fig. S1). In concert with the chemical reaction, a helix overlying the active site is partially melted into a β -hairpin-containing loop that adopts a closed conformation. Rearrangement from the open to closed conformation precisely locates U34 within the enzyme's active site. The structures have revealed the catalytic Cys residues, and suggest that the catalytic mechanism proceeds through persulphide chemistry.

Overall structure of the MnmA-tRNA complex

The *E. coli* MnmA and tRNA^{Glu} binary complex was co-crystallized in two different crystal forms—forms I and II (ref. 13). The form II structure was solved by multiwavelength anomalous dispersion (Supplementary Table S1 and Supplementary Fig. S2). The form I structure was solved by molecular replacement, using the form II structure as a search model (Supplementary Table S1 and Supplementary Fig. S3).

MnmA adopts a cradle-like shape (Fig. 1a, b) consisting of the amino-terminal catalytic domain (residues 4–215), the central domain (residues 216–277), and the carboxy-terminal β -barrel domain (residues 278–368). The catalytic domain is a mixed α/β -fold with five-stranded parallel and two-stranded antiparallel β -sheets, flanked by ten α -helices and a 'variable segment' (residues 187–215) (Fig. 1a). The catalytic domain shares structural similarity with the N-type ATP pyrophosphatase^{14,15}, and contains the P-loop motif 13-SGGVDSS-19 (Supplementary Fig. S4). The central domain is a four-stranded β -sheet with a long loop region, and the β -barrel domain resembles the C-terminal domain of EF-Tu¹⁶ (Fig. 1a).

Trapping the adenylated intermediate

Reaction of MnmA with tRNA^{Glu} and ATP led to the detection of an adenylated tRNA intermediate (Fig. 1c), consistent with the structural similarity to the N-type ATP pyrophosphatase, which

¹Department of Biological Information, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama-shi, Kanagawa 226-8501, Japan. ²Department of Chemistry and Biotechnology, Graduate School of Engineering, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. ³PRESTO, JST, Honcho, Kawaguchi-shi, Saitama 332-0012, Japan. ⁴RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan.

also activates its substrate as an adenylated intermediate^{14,15}. To investigate the reaction mechanism, the ternary complex including ATP was co-crystallized (form III crystal in ref. 13). The form III structure was determined by molecular replacement, using the form II structure as the search model (Supplementary Table S1). After simulated annealing refinement, we observed a clear electron density beside U34 in the $F_o - F_c$ map (Fig. 1d), which was interpreted as an adenylyl moiety. This is the first visualization of an adenylated RNA intermediate. We observed an additional electron density near the P α atom of the adenylated intermediate, interpreted as a sulphate ion from the mother liquor, mimicking the liberated pyrophosphate group.

Minor-groove interaction

The three domains of MnmA are complementary to the L-shaped tRNA, contacting the anticodon arm and D stem (Figs 1a, b and 2). The bottoms of the D and anticodon stems predominantly interact with the long loop of the central domain (Fig. 2a, b). The Gly-rich region (residues 245–252) in the long loop adopts an extended structure and horizontally invades the minor groove formed between the D and anticodon stems, where extensive van der Waals and hydrogen-bonding interactions form between the MnmA main chain and the RNA backbone (Fig. 2a, b). This novel minor-groove

interaction is further stabilized by the side chain of Arg 243, which forms salt bridges with the phosphate groups of A26 and G39 in the major groove (Fig. 2b). Although the side-chain carboxyl group of Glu 256 hydrogen-bonds to the N2 atom of G23, replacement of this position suggested a minor contribution to tRNA recognition (Fig. 2d).

Anticodon recognition in the form I structure

MnmA incorporates s²U into tRNA species with the anticodons UUC (glutamate), UUU (lysine) and UUG (glutamine), and, as expected, recognizes U34 and U35 in a base-specific manner (Fig. 2c). U34 is flipped out to enter the active site where it is recognized by the strictly conserved Gln 151, whose O ϵ and N ϵ atoms hydrogen-bond with the N3 and O2 atoms of U34, respectively (Fig. 2b, c). The Q151E mutant lost sulphur incorporation activity (Fig. 2d). U34 is further stabilized by an edge-to-face interaction with His 128 (Fig. 2c), and by hydrogen bonding between the 2'-hydroxyl group and the side chain of Asp 99 (Fig. 3d). The Lys 149 ϵ -amino group hydrogen-bonds to the U33 phosphate group (Supplementary Fig. S5), in agreement with the loss of activity in K149A (Fig. 2d), suggesting that the interaction is essential for properly anchoring U34. U35 is recognized by the strictly conserved Gln 344, whose O ϵ and N ϵ atoms hydrogen-bond with the N3 and O2 atoms of U35, respectively (Fig. 2c); Q344A is inactive (Fig. 2d). The O4 atom of U35 hydrogen-bonds with the main-chain amide group of Tyr 312 (Fig. 2c), and the base also contacts Asn 97 and Phe 154, probably forming a platform necessary for the proper location of U35 (Fig. 2b, c). N97A, but not F154A, drastically reduced sulphur incorporation, confirming the importance of this residue (Fig. 2d). The O2 atom of C36 forms a hydrogen bond with the guanidino group of the strictly conserved Arg 311 (Fig. 2c and Supplementary Fig. S4). R311A reduced sulphur incorporation sevenfold (Fig. 2d), suggesting that Arg311 fixes C36 to properly locate U34 in the active site. In addition to the interaction with the O2 atom of C36 in tRNA^{Glu}, Arg 311 could readily interact with either the O2 atom of U36 or the O6 atom of G36 in tRNA^{Lys} and tRNA^{Gln} (the other substrates for MnmA), respectively. Mutation of Thr 239, which also contacts C36, reduced sulphur incorporation activity twofold (Fig. 2d). The overall mechanism of tRNA recognition by MnmA is shown in Supplementary Fig. S5.

Structural rearrangement between form I and form II structures

Although most of the tRNA interactions in the form I structure are conserved in form II, notable differences are found around the active site and the 'variable segment' carrying the catalytic residue Cys 199. Note that the 'variable segment' is not involved in crystal packing in form I or form II. In form I, the 'variable segment' folds into the long helix Hv (residues 198–211), following a disordered region (residues 190–195) (Fig. 3a and Supplementary Fig. S6). In contrast, in form II, the N-terminal part of Hv is unfolded and restructured, together with the disordered region, into a loop (Lv; residues 187–195 and 204–205), a protruding β -hairpin (β v; residues 196–203), and a short, tilted helix (Fig. 3b and Supplementary Fig. S6). Accompanied by the structural rearrangement, U34 shifts towards the catalytic site by 2 Å and rotates by 140°, relative to form I (Fig. 3d, e). The newly formed hydrogen bond between the O4 atom of U34 and the His 128 side chain may attract the flipped base into the catalytic centre (Fig. 3e). The H128A replacement reduced sulphur incorporation tenfold (Fig. 2d). Hydrogen bonds between the main-chain carbonyl and amide groups of Gly 197 in β v with the 2'-hydroxyl group of U34 and the backbone phosphate of U35, respectively, are critical for the movement of U34 (data not shown). Concerted movement of Cys 199 appears to place U34 in the proper orientation by shrinking the active site pocket (Fig. 3e) and rotating the base by 140°. Repositioned U34 is stabilized by edge-to-face stacking interactions with Phe 154 instead of His 128, and by van der Waals interactions with Asp 99, Gln 151 and Cys 199 (Fig. 3d, e).

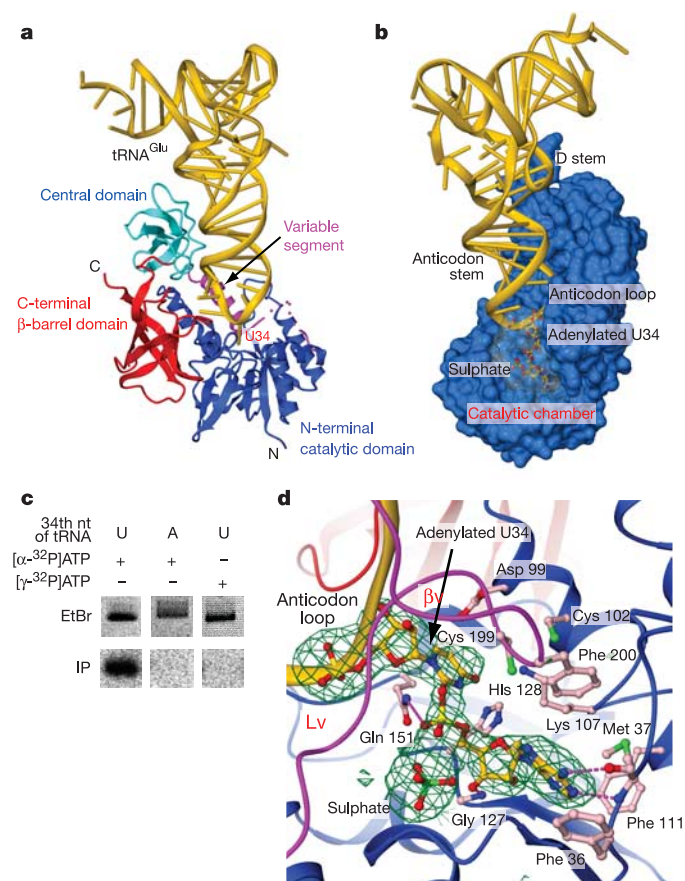


Figure 1 | Structure of the MnmA-tRNA^{Glu} complex. **a**, Form I complex structure. **b**, Overall interactions between MnmA (blue) and tRNA (yellow) in the adenylated intermediate form (form III). Residues 25–33, 51–53, 64–72, 207–213 and 218 of MnmA are shown in the translucent surface to highlight the catalytic chamber. **c**, Detection of the adenylated tRNA intermediate. tRNA^{Glu} with U34 (intact) or A34 (mutant) was incubated with $[\alpha\text{-}^{32}\text{P}]\text{ATP}$ or $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ (Supplementary Methods), and the products detected by ethidium bromide (EtBr) staining and on a phosphor imaging plate (IP). **d**, The σ_A -weighted $F_o - F_c$ map (2.5 σ) for adenylated U34. The carbon atoms of tRNA and MnmA are coloured yellow and pink, respectively.

Adenylated intermediate form (form III)

The form III structure is similar to form II, except that U34 is adenylated and rotated by $\sim 30^\circ$ (Fig. 3e, f). The adenyl moiety binds the hydrophobic cavity formed by the well-conserved residues Phe 36, Met 37, Lys 107, Phe 111 and Phe 200 (Fig. 1d). The main-chain amide and carbonyl groups of Met 37 hydrogen-bond with the adenine ring N1 and N6 atoms, respectively (Fig. 1d). The Gln 151 side chain hydrogen-bonds with the α -phosphate group of the adenyl moiety, and the main-chain amide of Gly 127 hydrogen-bonds with the 3'-hydroxyl group (Fig. 1d). F36A, M37A and F200A reduced sulphur incorporation 1.5-, 2.6- and 2.6-fold, respectively, relative to wild type (Fig. 2d), showing the importance of the location of the adenyl moiety for sulphur incorporation. The structure suggests the P-loop motif (Fig. 3f) functions in the adenylation of U34, as proposed for other N-type ATP pyrophosphatases^{14,15}. The D17A replacement abolished sulphuration activity (Fig. 2d), suggesting that Asp 17 helps liberate pyrophosphate from ATP.

Sequential snapshots of three reaction stages

The three crystal structures of the MnmA-tRNA^{Glu} complex are likely to represent snapshots of sequential steps in tRNA thiolation (Fig. 3); the initial tRNA-binding state (form I), the prereaction state (form II), and the adenylated intermediate state (form III). In the initial tRNA-binding state, the 'variable segment' folds into Hv and the residual disordered region, and the active site adopts an 'open'

conformation, which might be advantageous for initial binding of U34 (Fig. 3a). The 'open' structure is predominantly stabilized by hydrophobic interactions of Hv with H3a and H3b (Fig. 3a), and hydrogen bonds between Glu 203 and Arg 207 in Hv and the 2'-hydroxyl group of U35 (Fig. 2c). As the O2 thiolation site of U34 is specifically recognized by Gln 151, this initial binding state is inactive with respect to the adenylation and sulphuration reactions. The putative catalytic residues Cys 102 and Cys 199 (see below) form an intramolecular disulphide bond (Fig. 3d), reflecting the termination of sulphur incorporation in the previous reaction cycle (see Fig. 5). Comparable thermal factors for the S γ atoms of these two Cys residues ($65.3 \text{ \AA}^2$ and $73.1 \text{ \AA}^2$ for Cys 102 and Cys 199, respectively) support disulphide bond formation.

In the prereaction state, the helical variable segment is transformed to a continuous structure comprising Lv, β v and a short helix (Fig. 3b), and the active site adopts a 'closed' conformation with the catalytic pocket covered by β v and part of Lv (Fig. 3e). This rearrangement results from reorganization of the hydrophobic interactions of the 'variable segment' with H3a and H3b (Fig. 3a, b). The 'open' to 'closed' rearrangement relocates U34 of tRNA^{Glu} close to the catalytic centre with the base rotated (Fig. 3e). Consequently, the O2 atom of U34 no longer interacts with MnmA and is in a productive conformation, allowing it to attack the α -phosphate group of ATP leading to adenylation (Fig. 3e).

The adenylated intermediate form retains the 'closed' conformation

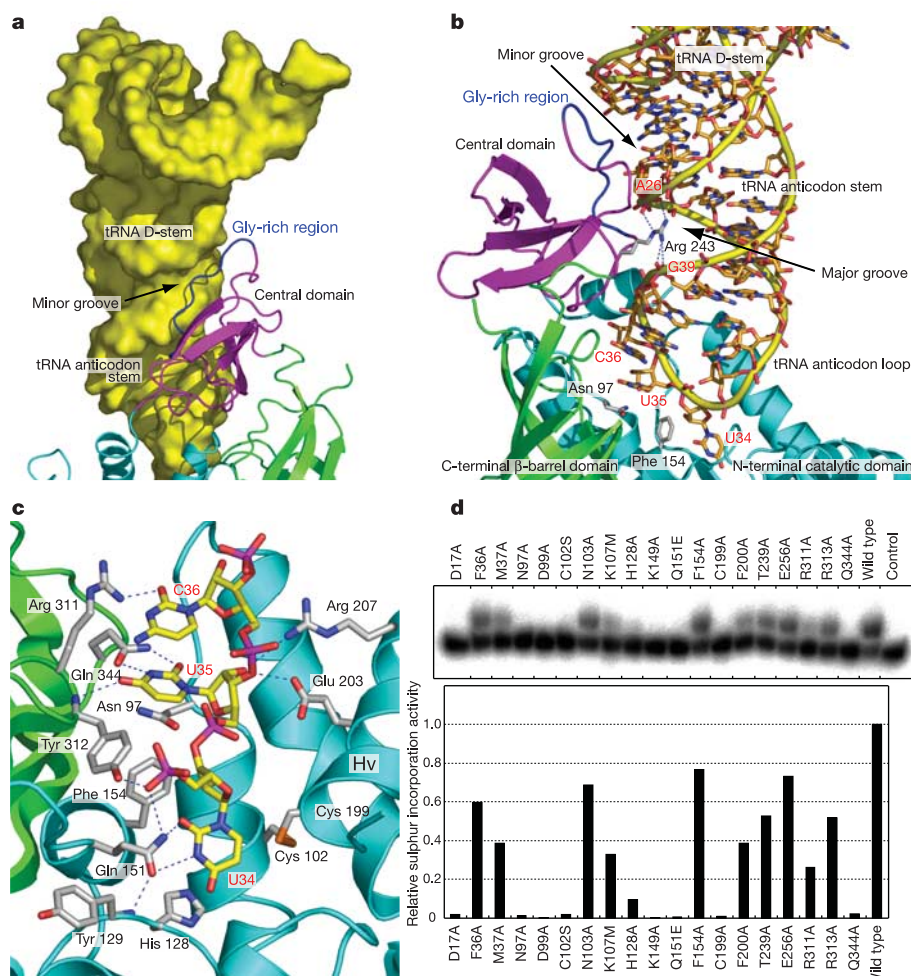


Figure 2 | tRNA recognition. **a**, Recognition of the D and anticodon stems by the Gly-rich loop (blue tube) in the central domain. **b**, Close-up view of tRNA recognition. **c**, Specific recognition of the anticodon triplet (yellow) in the form I structure. Hydrogen bonding (dashed lines) was inferred

specifying that the bond length is from 2.2 to 3.5 Å, and that the bond angle is from 160 to 200° in the case of the sp^2 hybrid orbital. The catalytic Cys residues, Cys 102 and Cys 199, form a disulphide bond in this state. **d**, *In vitro* sulphur incorporation analysis by gel retardation of MnmA mutants.

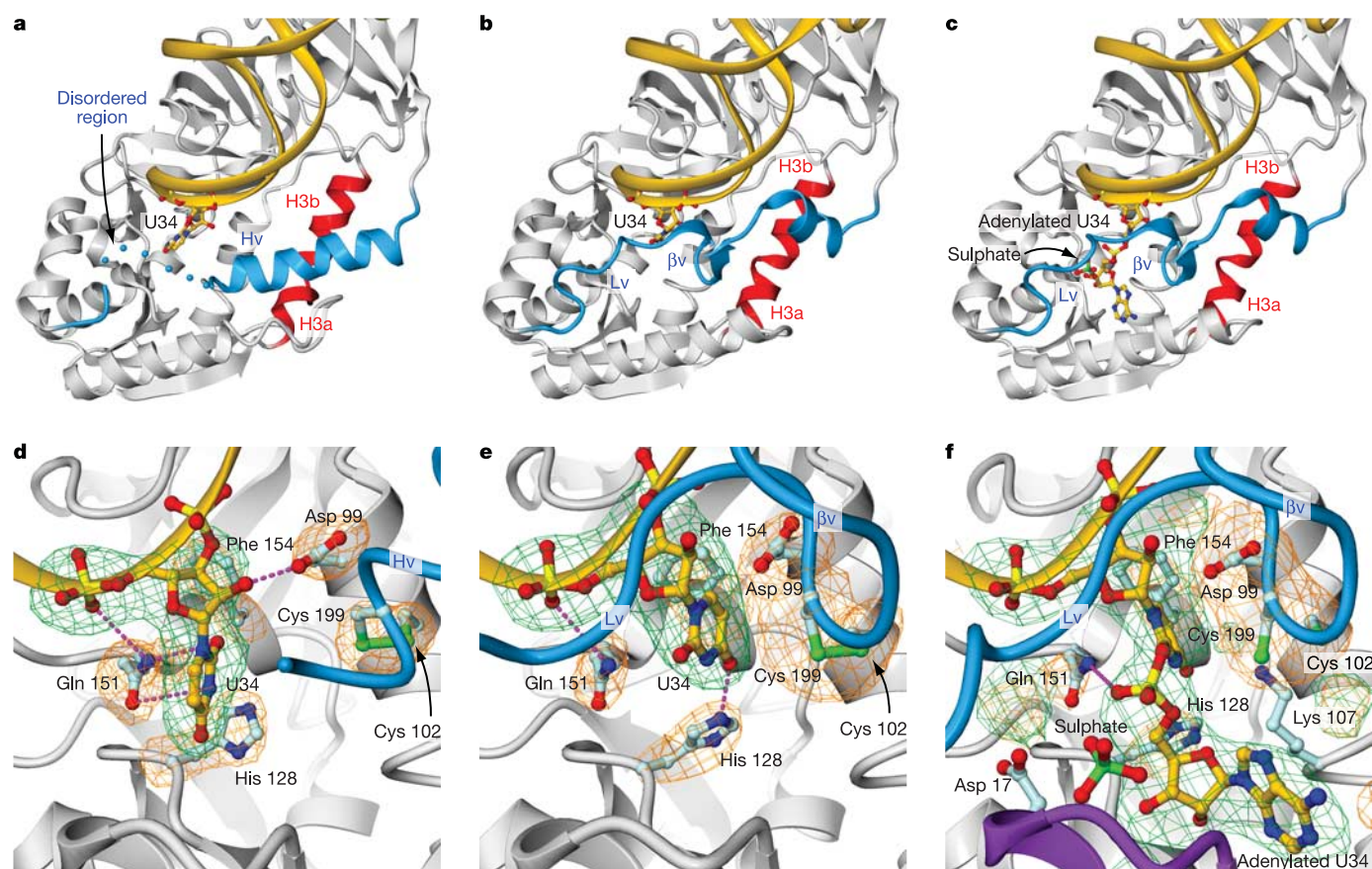


Figure 3 | Sequential snapshots of the sulphuration reaction. **a, d,** The initial tRNA-binding state. **b, e,** The prereaction state. **c, f,** The adenylated intermediate state. **a–c,** The overall conformational change of MnmA. The ‘variable segment’ is shown as a blue ribbon model. **d–f,** Close-up views of

the active site with F_o-F_c maps contoured at 3.5, 3.5 and 2.7 σ for MnmA residues (orange) and (adenylated)U34 (green) in forms I, II and III, respectively. The P-loop motif is coloured violet. Hv or β v is depicted by a tube model to clearly present the catalytic Cys 102 and Cys 199.

(Fig. 3c) of the prereaction form, except that the reduced thiol groups of Cys 102 and Cys 199 indicate an activated state (Fig. 3f). This is supported by the substantially different thermal factors of the S γ atoms (85.4 Å² and 58.2 Å² for Cys 102 and Cys 199, respectively, in molecule B). In the prereaction and adenylated intermediate forms, Lv and β v function as a lid on the active site, forming a catalytic chamber critical for excluding solvent during incorporation of the reactive sulphur at U34 (Figs 1b and 3). In the adenylated intermediate state, U34 rotates $\sim 30^\circ$ further, placing O2 in the vicinity of the catalytic Cys 199 (Fig. 3f). Thus, the adenylated U34 adopts a more productive conformation for sulphur acceptance from Cys 199.

Catalytic mechanism

MnmA catalyses the thiolation reaction in two steps (see Fig. 5): MnmA first activates the C2 atom of U34 by adenylation using ATP, and then the persulphide sulphur on the catalytic Cys residue is directly or indirectly transferred to the C2 carbon of U34, assisted by acid–base chemistry. The present structures of the MnmA–tRNA^{Glu} complex unambiguously defined Cys 102 and Cys 199 as catalytic residues critical for sulphur transfer to U34 (Figs 3 and 5). This is consistent with the C102S and C199S mutants completely losing thiolation activity (Fig. 2d). In the adenylated intermediate state, Cys 199 is located near U34, with its S γ atom 4.8 Å away from the C2 atom of U34 (Figs 1d and 3f). Cys 102 is on the opposite side of the uracil base relative to Cys 199 (Figs 1d and 3f), strongly suggesting that Cys 199 accepts the sulphur as a persulphide group from the sulphur relay system, and the terminal sulphur of the persulphide is transferred to the C2 atom of U34, with assistance from Cys 102 (Figs 3f and 5). Labelling of MnmA C102S, but not C199S, with

[³⁵S]sulphur during *in vitro* sulphur transfer with the sulphur relay components (IscS, TusA, TusBCD and TusE) indicates direct sulphur transfer from TusE to Cys 199 (Fig. 4). Sulphur transfer required tRNA and ATP (Fig. 4), in agreement with the observation that the catalytic Cys residues are first activated on tRNA binding and

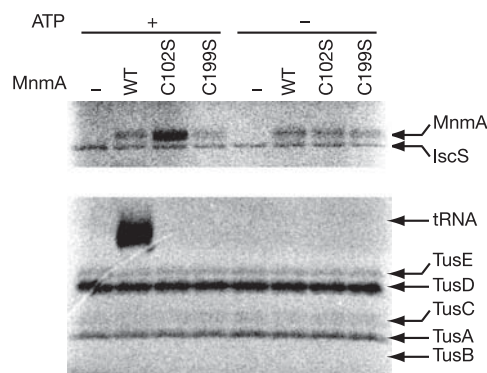


Figure 4 | *In vitro* sulphur transfer from IscS to tRNA^{Glu}. [³⁵S]-labelled sulphur retentions on wild-type (WT) MnmA, and C102S and C199S mutants, are analysed in the presence and absence of ATP. Although non-specific labelling is inevitable due to the intrinsic nature of the experiment, the band corresponding to the C102S mutant is much denser, whereas the band corresponding to C199S is much weaker, as compared with that for the wild type. In agreement with the *in vitro* sulphur incorporation analysis shown in Fig. 2d, tRNA was not labelled by the C102S or C199S mutants.

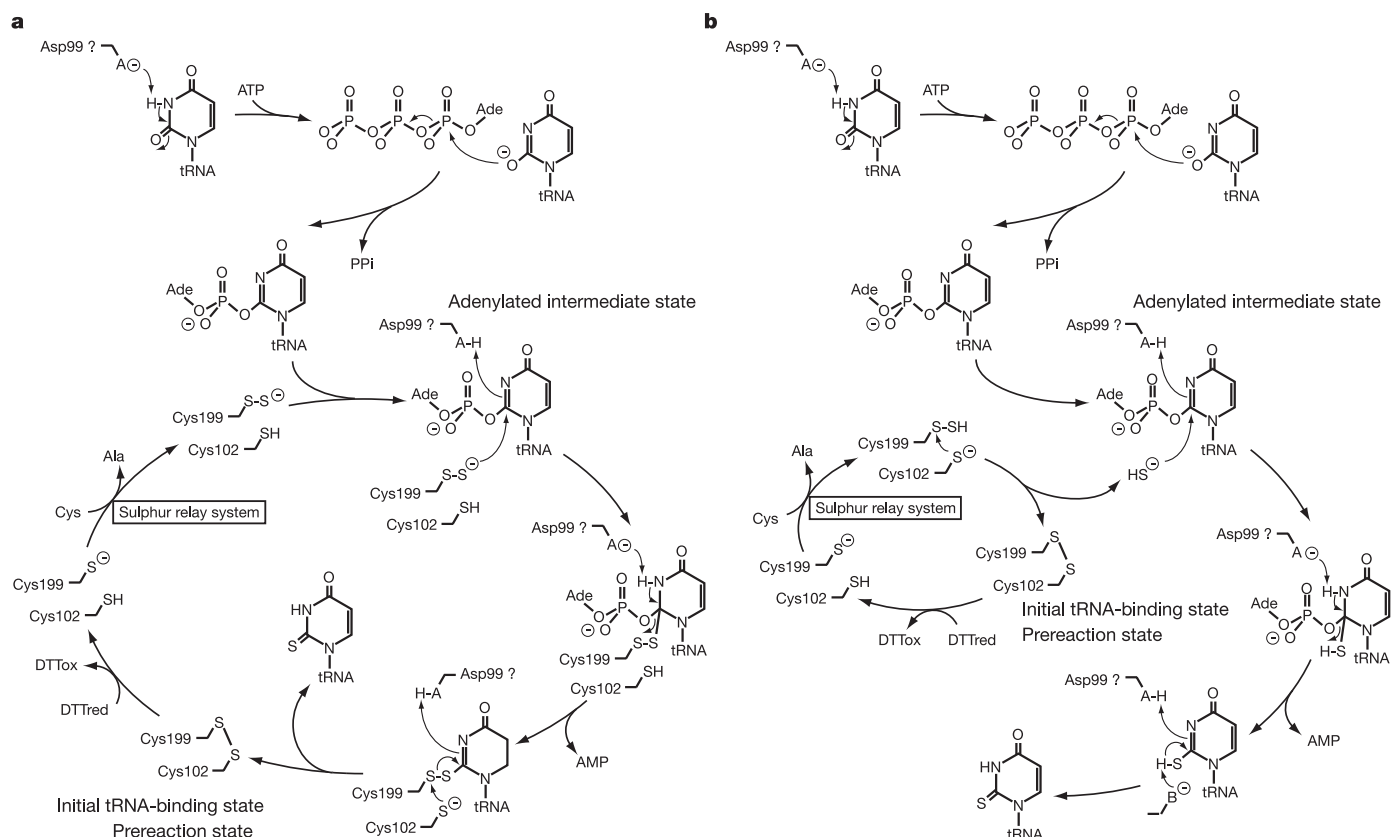


Figure 5 | Catalytic mechanisms of MnmA. **a, b**, Two possible schemes of an MnmA-catalysed reaction are presented, showing the direct sulphur transfer mechanism (**a**) and the hydrogen-sulphide-mediated mechanism (**b**).

adenylation. Two alternative schemes for the MnmA catalysed reaction can be proposed (Fig. 5), as suggested for the s⁴U modification^{17,18}. In the first scheme, the terminal sulphur of the persulphide on Cys199 directly attacks the activated uridine (Fig. 5a), and in the second scheme the terminal sulphur of the persulphide is liberated as hydrogen sulphide, which serves as a nucleophile to displace the activated oxygen of the uracil base (Fig. 5b). In the first scheme, but not the second, C102S would accumulate the covalently bonded MnmA–tRNA intermediate, which should be detected during the sulphur transfer experiment. The present data favour the second scheme, where hydrogen sulphide mediates sulphur transfer from MnmA to tRNA (Figs 4 and 5). In both cases, it is essential that the second catalytic cysteine, Cys 102, nucleophilically attacks the S_γ atom of the catalytic persulphide of Cys199 to form a covalent linkage^{17,18} (Fig. 5). The initial tRNA-binding and prereaction form structures present the disulphide bond formation between Cys 102 and Cys 199 (Fig. 3d, e). The proposed catalytic mechanisms require general acid–base catalysts, which de-protonate and protonate the N3 atom of U34 during the reaction (Fig. 5). The present structure (Fig. 3f) combined with the sulphuration analysis (Fig. 2d) suggests that Asp 99 is the most plausible acid–base catalyst.

Discussion

Three discrete MnmA–tRNA^{Glu} structures provide snapshots of the chemical reaction in which dynamic structural rearrangements relocate the target U34 close to the catalytic site while activating two catalytic Cys residues in a reduced state. The resultant closed conformation of the complex facilitates accurate sulphur incorporation by forming a catalytic chamber to shield the active site from the solvent (Supplementary Fig. S1), which might be a general mechanism for an enzyme to incorporate a reactive atom at a precise position in a macromolecule.

METHODS

Structure determination and refinement. Form I, II and III crystals belong to the space groups C2, I2₁2₁ and C2, respectively¹³. An unmodified *in vitro* transcript of tRNA^{Glu} isoacceptor with UUC anticodon was used. First, the form II complex structure was determined by the multiple anomalous dispersion (MAD) method using selenomethionine-labelled MnmA¹³. Five selenium sites were determined with the program SnB¹⁹. The selenium parameters were refined and the phases were calculated at 3.5 Å resolution using the program SHARP²⁰, with an acentric figure of merit of 0.530. The initial phases were improved by density modification using SOLOMON²¹. The initial model was manually built using the program O²², and iterative cycles of refinement with CNS²³ and manual rebuilding with O²² were carried out at 3.4 Å resolution, until the *R*_{free} factor converged. The form I structure was solved at 3.1 Å resolution by molecular replacement with the program CNS²³, using the refined form II structure as a search model. The program DM²⁴ and CNS²³ was used for two-fold non-crystallographic symmetry (NCS) averaging and density modification. The form III structure was determined at 3.4 Å resolution by molecular replacement with the program MOLREP²⁴, using the form II structure as a search model. A summary of the final refinement statistics is available in Supplementary Table S1. The stereochemistry of the models, as assessed by PROCHECK²⁵, showed that 96.7%, 94.9% and 95.1% of the protein residues in the form I, II and III structures, respectively, are in the most favoured and additionally allowed regions of the Ramachandran plot. On the other hand, 0%, 0% and 0.2% of the residues in the form I, II and III crystals, respectively, are in the disallowed region. The four residues (0.2%) in the disallowed portion in the form III structure are located on loop regions and are not disordered. Molecular graphics were illustrated with PyMol²⁶ and CueMol²⁷.

Biochemical analyses. The details of biochemical analyses, such as detection of the adenylated intermediate, *in vitro* sulphur incorporation analysis, and *in vitro* sulphur transfer experiment, are described in Supplementary Information, and are also described in references 12, 28–30.

Received 6 April; accepted 12 May 2006.

1. Björk, G. R. Biosynthesis and function of modified nucleosides. In *tRNA: Structure, Biosynthesis and Function* (eds Söll, D. & RajBhandary, U. L.) 165–205 (American Society for Microbiology, Washington DC, 1995).

2. Yokoyama, S. & Nishimura, S. Modified nucleosides and codon recognition. In *tRNA: Structure, Biosynthesis, and Function* (eds Söll, D. & RajBhandary, U. L.) 207–224 (American Society for Microbiology, Washington DC, 1995).
3. Yokoyama, S. *et al.* Molecular mechanism of codon recognition by tRNA species with modified uridine in the first position of the anticodon. *Proc. Natl Acad. Sci. USA* **82**, 4905–4909 (1985).
4. Ashraf, S. S. *et al.* Single atom modification (O → S) of tRNA confers ribosome binding. *RNA* **5**, 188–194 (1999).
5. Krüger, M. K., Pedersen, S., Hagervall, T. G. & Sørensen, M. A. The modification of the wobble base of tRNA^{Glu} modulates the translation rate of glutamic acid codons *in vivo*. *J. Mol. Biol.* **284**, 621–631 (1998).
6. Urbonavicius, J., Qian, Q., Durand, J. M., Hagervall, T. G. & Björk, G. R. Improvement of reading frame maintenance is a common function for several tRNA modifications. *EMBO J.* **20**, 4863–4873 (2001).
7. Sylvers, L. A., Rogers, K. C., Shimizu, M., Ohtsuka, E. & Söll, D. A 2-thiouridine derivative in tRNA^{Glu} is a positive determinant for aminoacylation by *Escherichia coli* glutamyl-tRNA synthetase. *Biochemistry* **32**, 3836–3841 (1993).
8. Isel, C., Marquet, R., Keith, G., Ehresmann, C. & Ehresmann, B. Modified nucleotides of tRNA^{Lys} modulate primer/template loop–loop interaction in the initiation complex of HIV-1 reverse transcription. *J. Biol. Chem.* **268**, 25269–25272 (1993).
9. Isel, C. *et al.* Structural basis for the specificity of the initiation of HIV-1 reverse transcription. *EMBO J.* **18**, 1038–1048 (1999).
10. Sullivan, M. A., Cannon, J. F., Webb, F. H. & Bock, R. M. Antisuppressor mutation in *Escherichia coli* defective in biosynthesis of 5-methylaminomethyl-2-thiouridine. *J. Bacteriol.* **161**, 368–376 (1985).
11. Kambampati, R. & Lauhon, C. T. MnmA and LscS are required for *in vitro* 2-thiouridine biosynthesis in *Escherichia coli*. *Biochemistry* **42**, 1109–1117 (2003).
12. Ikeuchi, Y., Shigi, N., Kato, J., Nishimura, A. & Suzuki, T. Mechanistic insights into sulfur relay by multiple sulfur mediators involved in thiouridine biosynthesis at tRNA wobble positions. *Mol. Cell* **21**, 97–108 (2006).
13. Numata, T. *et al.* Crystallization and preliminary X-ray analysis of the tRNA thiolation enzyme MnmA from *Escherichia coli* complexed with tRNA^{Glu}. *Acta Crystallogr. F* **62**, 368–371 (2006).
14. Tesmer, J. J., Klem, T. J., Deras, M. L., Davisson, V. J. & Smith, J. L. The crystal structure of GMP synthetase reveals a novel catalytic triad and is a structural paradigm for two enzyme families. *Nature Struct. Biol.* **3**, 74–86 (1996).
15. Rizzi, M. *et al.* Crystal structure of NH₃-dependent NAD⁺ synthetase from *Bacillus subtilis*. *EMBO J.* **15**, 5125–5134 (1996).
16. Nissen, P. *et al.* Crystal structure of the ternary complex of Phe-tRNA^{Phe}, EF-Tu, and a GTP analog. *Science* **270**, 1464–1472 (1995).
17. Palenchar, P. M., Buck, C. J., Cheng, H., Larson, T. J. & Mueller, E. G. Evidence that Thil, an enzyme shared between thiamin and 4-thiouridine biosynthesis, may be a sulfurtransferase that proceeds through a persulfide intermediate. *J. Biol. Chem.* **275**, 8283–8286 (2000).
18. Mueller, E. G., Palenchar, P. M. & Buck, C. J. The role of the cysteine residues of Thil in the generation of 4-thiouridine in tRNA. *J. Biol. Chem.* **276**, 33588–33595 (2001).
19. Weeks, C. M. & Miller, R. The design and implementation of SnB version 2.0. *J. Appl. Cryst.* **32**, 120–124 (1999).
20. de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
21. Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F₁ ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
22. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
23. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
24. Collaborative Computational Project Number 4. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
25. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
26. Delano, W. L. The PyMOL molecular graphics system. v.0.97 (<http://www.pymol.org>) (2002).
27. Ishitani, R. CueMol: Molecular visualization framework. v.1.1 (<http://cuemol.sourceforge.jp>) (2006).
28. Ikeuchi, Y. *et al.* Molecular mechanism of lysidine synthesis that determines tRNA identity and codon recognition. *Mol. Cell* **19**, 235–246 (2005).
29. Shigi, N., Suzuki, T., Tamakoshi, M., Oshima, T. & Watanabe, K. Conserved bases in the TPsi C loop of tRNA are determinants for thermophile-specific 2-thiouridylation at position 54. *J. Biol. Chem.* **277**, 39128–39135 (2002).
30. Shigi, N. *et al.* Temperature-dependent biosynthesis of 2-thioribothymidine of *Thermus thermophilus* tRNA. *J. Biol. Chem.* **281**, 2104–2113 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature. A figure summarizing the main result of the paper is also included.

Acknowledgements We thank M. Ibba for helpful improvement of the manuscript. We thank M. Kawamoto and N. Shimizu for their help in data collection at SPring-8. This work was supported by a PRESTO program grant from Japan Science and Technology (JST) to O.N., a grant for the National Project on Protein Structural and Functional Analyses from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) to O.N., grants from MEXT to O.N., S.F. and T.S., a JSPS Fellowship for Japanese Junior Scientists to Y.I., the Asahi Glass Foundation to S.F., and Sumitomo and Kurata Memorial Hitachi Science and Technology Foundation grants to O.N.

Author Contributions T.N. carried out structural determination and mutant preparation, and wrote the paper with editing from S.F. and O.N. S.F. and O.N. assisted the structural determination. Y.I. carried out the biochemical analyses under supervision of T.S. All authors discussed the results and commented on the manuscript. O.N. supervised all of the work.

Author Information Coordinates and structure factors are deposited in the Protein Data Bank under accession code 2DER, 2DET and 2DEU for the initial tRNA-binding form, the prereaction form, and the adenylated intermediate form, respectively. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to O.N. (nureki@bio.titech.ac.jp).

A low fraction of nitrogen in molecular form in a dark cloud

S. Maret¹, E. A. Bergin¹ & C. J. Lada²

Nitrogen is the fifth most abundant element in the Universe. In the interstellar medium, it has been thought to be mostly molecular (N_2)¹. However, N_2 has no observable rotational or vibrational transitions, so its abundance in the interstellar medium remains poorly known. In comets, the N_2 abundance is very low^{2,3}, while the elemental nitrogen abundance is deficient with respect to the solar value. Moreover, large nitrogen isotopic anomalies are observed in meteorites and interstellar dust particles⁴. Here we report the N_2H^+ (and by inference the N_2) abundance inside a cold dark molecular cloud. We find that only a small fraction of nitrogen in the gas phase is molecular, with most of it being atomic. Because the compositions of comets probably reflect those of dark clouds⁵, this result explains the low N_2 abundance in comets. We argue that the elemental nitrogen abundance deficiency in comets can be understood if the atomic oxygen abundance is lower than predicted by present chemical models. Furthermore, the lack of molecular nitrogen in molecular clouds explains the nitrogen anomalies in meteorites and interstellar dust particles, as nitrogen fractionation is enhanced if gaseous nitrogen is atomic⁶.

For more than three decades, chemical models¹ have predicted that nitrogen is mostly in the form of N_2 in dense interstellar clouds. Observations of N_2H^+ —a chemical daughter product of N_2 —have been used, together with HCO^+ observations, to estimate the N_2/CO abundance ratio in molecular clouds⁷. In these clouds, both ions are formed from the reaction with the H_3^+ ion with either N_2 or CO , and are destroyed mostly by recombination with electrons. Both reactions occur at a similar rate for the two ions, and so the measure of the HCO^+/N_2H^+ abundance ratio is a direct measure of the N_2/CO ratio, and in turn of the N_2 abundance if the CO abundance is known. Observations of this ratio indicate that most nitrogen is in the form of N_2 in molecular clouds, thereby supporting theoretical predictions.

This approach is too simple, omitting consideration of the many other formation and destruction routes of N_2H^+ that are known to exist in interstellar clouds. For example, N_2H^+ is destroyed via a reaction with CO that has no counterpart for HCO^+ . Therefore, to derive the N_2 abundance from N_2H^+ , one must consider an expanded network of gas-phase chemical reactions. Furthermore, in dense and cold parts of molecular clouds, molecules can condense on interstellar grains and form an icy mantle around them, a phenomenon known as depletion. Depletion of several molecules is observed in numerous cold molecular cores⁸. For example, CO and CS abundances are estimated to decrease by several orders of magnitude near the core centres, where the density is highest⁹. Although N_2H^+ itself does not deplete on grain mantles, its abundance also seems to be reduced in the centre regions of starless cores with respect to the core edges^{9,10}.

This disappearance from the gas phase is probably due to the

freeze-out of N_2H^+ parent molecule, N_2 , onto the grain surfaces. The ability of a molecule or atom to remain on a grain once it has frozen out depends on its binding energy to the grain surface. Observations of $C^{18}O$ and N_2H^+ emission are well reproduced by gas-phase chemistry models including depletion, provided that the ratio between the binding energies of N_2 and CO is between 0.5 and 0.75 (ref. 9). However, recent laboratory measurements indicate that

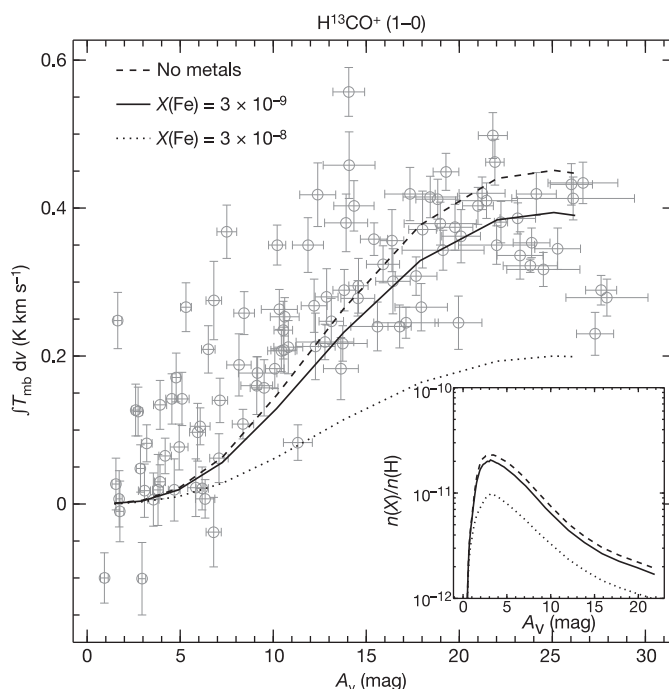


Figure 1 | Determination of the metal abundance inside the B68 core. Grey points with error bars (1σ) represent the observed $H^{13}CO^+$ (1–0) line intensities in main beam temperature units (T_{mb}) as a function of the visual extinction in the core (A_v). The black lines represent the model predictions for different metal abundances. The inset shows the corresponding $H^{13}CO^+$ abundances (X). The $H^{13}CO^+$ abundances have been computed as a function of the visual extinction by a chemical network that includes depletion and desorption (the removal of molecules from grain surfaces). The obtained abundances profile has been used to compute the line emission of the core as a function of the visual extinction using a Monte Carlo radiative transfer code. Finally, the model predictions have been compared to observations. The best agreement between the model predictions and the observations is obtained for a metal abundance of 3×10^{-9} with respect to H nuclei, but the observations are also consistent with a complete depletion of metals on the grain surfaces (that is, with the cosmic rays as the sole source of ionization).

¹Department of Astronomy, University of Michigan, 500 Church Street, Ann Arbor, Michigan 48109-1042, USA. ²Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.

the binding energies of these two molecules are almost the same¹¹, casting the previous interpretation into doubt. Besides, other laboratory measurements¹² of the dissociative recombination rate of N_2H^+ —a key parameter to derive N_2 abundance from observations of N_2H^+ —shows that this reaction, which was previously assumed to form N_2 , instead mostly forms NH . In view of these recent results, the abundance of N_2 in molecular clouds has to be re-investigated.

Using a technique that combines the predictions of chemical network¹³ with a radiative transfer model, we have re-analysed the N_2H^+ emission maps in the B68 pre-stellar core. This core is an ideal laboratory to study interstellar chemistry, because its physical structure (density and temperature) is well known. The density profile of this core has been inferred from the extinction of background starlight¹⁴, and is found to be well described by the equations for a pressure-confined, self-gravitating isothermal sphere that is critically stable. In addition, the gas temperature inside the core has been constrained by comparison of a chemical/physical model to observations of $C^{18}O$ and ^{13}CO line emission¹⁵.

To model the N_2H^+ emission, it is crucial to constrain the abundance of electrons and CO , which set the destruction rate of this ion. The CO abundance has been determined in the core by modelling and observations of transitions of CO isotopologues¹⁵. A ready source of ionization is the ultraviolet photons emitted by nearby stars, but these cannot penetrate the interior of dense molecular cores. Consequently, cosmic rays are believed to provide the main source of ionization. In addition, pre-existing gaseous refractory metals of low ionization potential (for example, Fe and Mg) can contribute to the electron abundance because they can exchange charge with molecular ions, producing Fe^+ and Mg^+ , which are destroyed relatively slowly by radiative recombination¹⁶. In our model, we have assumed a standard value for the cosmic ionization rate ($\zeta = 3 \times 10^{-17} \text{ s}^{-1}$) and we have determined the electron abundance from observations of $H^{13}CO^+$ (1–0) line emission (see Fig. 1). Figure 2 shows the predicted electron abundance structure in the core. Our estimates are in agreement with electron abundances obtained in other cores^{16,17}.

The electron and CO abundances determine the destruction rate of N_2H^+ . The cosmic ionization rate sets the H_3^+ abundance, and thus the N_2H^+ production rate from N_2 . With these two factors constrained, the N_2H^+ abundance depends solely on the N_2 abundance inside the core (see Supplementary Information). Assuming that all nitrogen is initially in the form of N_2 , we found that the predicted N_2H^+ (1–0) line fluxes are consistently higher than the

observed ones (see Fig. 3). Therefore, we suggest that the nitrogen element in cold molecular clouds is mostly atomic and not molecular, as previously thought. These observations require only a quarter of the elemental nitrogen to be initially in the form of N . Indeed, earlier N_2H^+ observations did seem to support this idea¹⁸.

This result has important implications on our understanding of the composition of comets. Comets are thought to be, in whole or in part, aggregates of interstellar grains that survived the formation of the solar nebula⁵. The striking similarities between the composition of cometary and interstellar ices¹⁹ favour this hypothesis. However, some differences exist: the total amount of nitrogen is known to be deficient in comets, with respect to the solar value²⁰. This deficiency has been linked to the low abundance of N_2 in cometary ices^{2,3}. This low abundance is puzzling, if nitrogen is mostly in the form of N_2 in cold dark cloud as has long been assumed. However, it has previously been suggested that the low N_2 abundance in comets could be a result of a high N/N_2 ratio⁶. Indeed, our work indicates that only a small fraction of nitrogen is in form of N_2 in the gas phase (see Fig. 2). Therefore, the amount of N_2 that freezes on the grain surfaces is also small with respect to the total nitrogen abundance. Our model predicts that the N_2 abundance in grain mantles with respect to water will be only about 1%. Thus the lack of molecular nitrogen in cometary nuclei can now be understood.

The overall nitrogen deficiency in comets is more difficult to understand. In principle, if nitrogen is mostly in atomic form in the gas phase, only a fraction of the total nitrogen would be incorporated in grain mantles, because atomic nitrogen has a very low binding

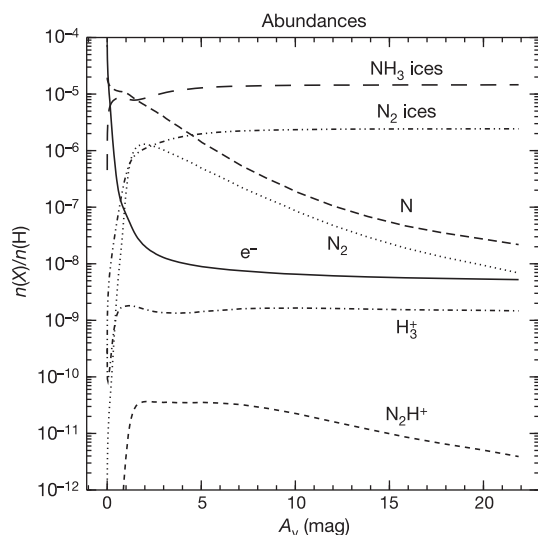


Figure 2 | Abundances of electrons and nitrogen-bearing species inside the core, at $t = 10^5$ yr.

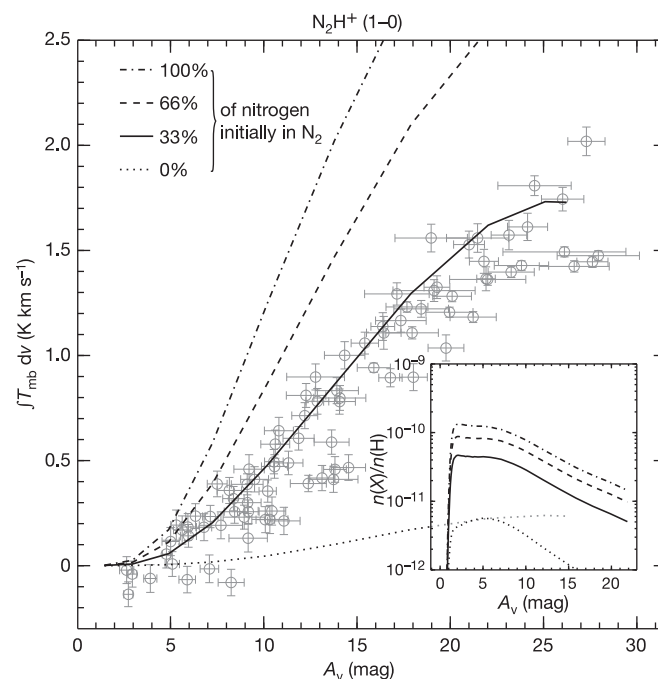


Figure 3 | Determination of the fraction of the elemental nitrogen in molecular form. Grey points with error bars (1σ) represent the observed N_2H^+ (1–0) line intensities as a function of A_v (ref. 9). The black lines represent the model predictions for different fractions of the nitrogen element initially in the form of N_2 . We refer to initial abundances as the abundances at $t = 0$ in our time-dependent chemical modelling (see Supplementary Information for further details). We have adopted an elemental nitrogen abundance of 2×10^{-5} with respect to H . N_2H^+ has an hyperfine structure owing to the interaction of the molecular electric field gradient with the electric quadrupole moments of the two nitrogen nuclei²⁴. This interaction splits the 1–0 line into seven components that must be modelled properly to obtain a correct opacity for the line emission. We have used a recent technique to handle the hyperfine structure within a Monte Carlo radiative transfer code²⁵, and properly computed the emission.

energy, and therefore does not condense on the grain surface. However, our model and observations require atomic nitrogen to form ammonia on grain mantles by grain surface reaction, in order to prevent it returning to the gas phase and forming N_2 . In the B68 prestellar core, our model indicates that the main nitrogen reservoir is ammonia ice, with an abundance of about 7% with respect to water. Although this value is compatible with ammonia ice abundances measured in dark clouds²¹, it is much higher than the ammonia abundances observed in comets ($\sim 1\%$)³. Our work could be reconciled with the nitrogen deficiency in comets if the atomic oxygen in the gas phase were lower than the one predicted by our model. This would prevent the formation of N_2 through NO, providing more atomic nitrogen in the gas phase and less NH_3 (see Supplementary Information).

This result could also explain the nitrogen isotopes anomalies ($^{15}N/^{14}N$) observed in some meteorites and interplanetary dust particles⁴. ^{15}N enrichments are probably the result of low-temperature chemistry, although dark cloud chemical models predict only modest enhancements of the $^{15}N/^{14}N$ ratio²². However, this fractionation is greatly increased if a large portion of the nitrogen is in atomic form⁶.

Finally, this work complements the recent detection of the N_2 in a diffuse molecular cloud²³. The N_2 abundance is expected to be much lower than in dense molecular clouds, because strong ultraviolet fields in diffuse molecular clouds photo-dissociate N_2 . The abundance measured in a diffuse molecular cloud (1.6×10^{-7} with respect to H nuclei) is about an order of magnitude lower than the one estimated in the present work (1.1×10^{-6} , see Fig. 2), and is therefore consistent with this picture.

In conclusion, this study emphasizes the similarity between the compositions of grain ices and of cometary nuclei, meteorites, and interplanetary dust particles. *In situ* analysis of cometary nuclei will be needed to confirm our hypothesis. The ESA's Rosetta mission—which will land on the comet 67P/Churyumov-Gerasimenko—will provide new insights into the nucleus composition of the comet, and any potential link to interstellar chemistry.

Received 27 February; accepted 16 May 2006.

- Herbst, E. & Klemperer, W. The formation and depletion of molecules in dense interstellar clouds. *Astrophys. J.* **185**, 505–534 (1973).
- Cochran, A. L., Cochran, W. D. & Barker, E. S. N_2^+ and CO^+ in Comets 122P/1995 S1 (deVico) and C/1995 O1 (Hale-Bopp). *Icarus* **146**, 583–593 (2000).
- Iro, N., Gautier, D., Hersant, F., Bockelée-Morvan, D. & Lunine, J. I. An interpretation of the nitrogen deficiency in comets. *Icarus* **161**, 511–532 (2003).
- Messenger, S., Keller, L. P., Thomas, K. L. & Walker, R. M. Nitrogen petrography in two ^{15}N -rich IDPs. *Meteorit. Planet. Sci.* **31**, A88 (1996).
- Irvine, W. M., Schloerb, F. P., Crovisier, J., Fegley, B. & Mumma, M. J. Comets: a link between interstellar and nebular chemistry. *Protostars Planets IV*, 1159–1200 (2000).
- Charnley, S. B. & Rodgers, S. D. The end of interstellar chemistry as the origin of nitrogen in comets and meteorites. *Astrophys. J. Lett.* **569**, L133–L137 (2002).
- Linke, R. A., Langer, W. D. & Guelin, M. Detection of $H^{15}NN^+$ and $HN^{15}N^+$ in interstellar clouds. *Astrophys. J. Lett.* **271**, L85–L88 (1983).
- Tafalla, M., Myers, P. C., Caselli, P., Walmsley, C. M. & Comito, C. Systematic molecular differentiation in starless cores. *Astrophys. J.* **569**, 815–835 (2002).
- Bergin, E. A., Alves, J., Huard, T. & Lada, C. J. N_2H^+ and $C^{18}O$ depletion in a cold dark cloud. *Astrophys. J. Lett.* **570**, L101–L104 (2002).
- Pagani, L., Pardo, J.-R., Apponi, A. J., Bacmann, A. & Cabrit, S. L183 (L134N) revisited. III. The gas depletion. *Astron. Astrophys.* **429**, 181–192 (2005).
- Öberg, K. I. et al. Competition between CO and N_2 desorption from interstellar ices. *Astrophys. J. Lett.* **621**, L33–L36 (2005).
- Geppert, W. D. et al. Dissociative recombination of N_2H^+ : Evidence for fracture of the NN bond. *Astrophys. J.* **609**, 459–464 (2004).
- Bergin, E. A., Langer, W. D. & Goldsmith, P. F. Gas-phase chemistry in dense interstellar clouds including grain surface molecular depletion and desorption. *Astrophys. J.* **441**, 222–243 (1995).
- Alves, J. F., Lada, C. J. & Lada, E. A. Internal structure of a cold dark molecular cloud inferred from the extinction of background starlight. *Nature* **409**, 159–161 (2001).
- Bergin, E. A. et al. The thermal structure of gas in pre-stellar cores: a case study of Barnard 68. *Astrophys. J.* (in the press).
- Caselli, P., Walmsley, C. M., Terzieva, R. & Herbst, E. The ionization fraction in dense cloud cores. *Astrophys. J.* **499**, 234–249 (1998).
- Caselli, P. et al. Molecular ions in L1544. II. The ionization degree. *Astrophys. J.* **565**, 344–358 (2002).
- Womack, M., Ziurys, L. M. & Wyckoff, S. Estimates of N_2 abundances in dense molecular clouds. *Astrophys. J.* **393**, 188–192 (1992).
- Bockelée-Morvan, D. et al. New molecules found in comet C/1995 O1 (Hale-Bopp). Investigating the link between cometary and interstellar material. *Astron. Astrophys.* **353**, 1101–1114 (2000).
- Wyckoff, S., Tegler, S. C. & Engel, L. Nitrogen abundance in comet Halley. *Astrophys. J.* **367**, 641–648 (1991).
- Gibb, E. L., Whittet, D. C. B., Boogert, A. C. A. & Tielens, A. G. G. M. Interstellar ice: the Infrared Space Observatory legacy. *Astrophys. J. Suppl.* **151**, 35–73 (2004).
- Terzieva, R. & Herbst, E. The possibility of nitrogen isotopic fractionation in interstellar clouds. *Mon. Not. R. Astron. Soc.* **317**, 563–568 (2000).
- Knauth, D. C., Andersson, B.-G., McCandless, S. R. & Warren Moos, H. The interstellar N_2 abundance towards HD 124314 from far-ultraviolet observations. *Nature* **429**, 636–638 (2004).
- Caselli, P., Myers, P. C. & Thaddeus, P. Radio-astronomical spectroscopy of the hyperfine structure of N_2H^+ . *Astrophys. J. Lett.* **455**, L77–L80 (1995).
- Keto, E., Rybicki, G. B., Bergin, E. A. & Plume, R. Radiative transfer and starless cores. *Astrophys. J.* **613**, 355–373 (2004).

Supplementary Information is linked to the online version of this paper at www.nature.com/nature.

Acknowledgements This work is supported by the National Science Foundation. S.M. is grateful to E. Herbst for discussions about the N_2 formation reaction rates. S.M. also thanks E. Keto for discussions about the radiative transfer of N_2H^+ hyperfine structure lines.

Author contributions S.M. and E.A.B. performed the chemical and radiative transfer modelling presented in this paper. S.M. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare that they have no competing financial interests. Correspondence and requests for materials should be addressed to S.M. (smaret@umich.edu).

LETTERS

Methane storms on Saturn's moon Titan

R. Hueso¹ & A. Sánchez-Lavega¹

The presence of dry fluvial river channels and the intense cloud activity in the south pole of Titan over the past few years^{1–3} suggest the presence of methane rain. The nitrogen atmosphere of Titan therefore appears to support a methane meteorological cycle that sculpts the surface and controls its properties^{1,4}. Titan and Earth are the only worlds in the Solar System where rain reaches the surface, although the atmospheric cycles of water and methane are expected to be very different⁵. Here we report three-dimensional dynamical calculations showing that severe methane convective storms accompanied by intense precipitation may occur in Titan under the right environmental conditions. The strongest storms grow when the methane relative humidity in the middle troposphere is above 80 per cent, producing updrafts with maximum velocities of 20 m s^{-1} , able to reach altitudes of 30 km before dissipating in 5–8 h. Raindrops of 1–5 mm in radius produce precipitation rainfalls on the surface as high as 110 kg m^{-2} and are comparable to flash flood events on Earth⁶.

Transient clouds in the lower atmosphere of Titan were first discovered from Earth-based observations^{7,8} and are in recent times a relatively common feature in the south pole of Titan and at 40° S latitude^{1,3,9–11}. There have been several proposals to explain the presence of these clouds, encompassing wet-methane convection^{12–14}, geological processes (specific for 40° S , ref. 10), and cloud formation by a seasonally controlled general circulation¹⁵. Methane convective storms can explain the spatial structure and rapid evolution of the thick clouds observed over the past years (during Titan's summer time) around Titan's south pole, and the severe rain accompanying the storms could explain the river channels seen on Titan's surface^{2,16}. However, little theoretical work has been done to study the cloud properties at a local scale and their associated precipitation. This is an essential task if we want to understand how the methane cycle works on Titan.

Our model works at the mesoscale (cloud lifetime of hours, spatial scales of a hundredth of a kilometre), and represents an advance over previous studies on convection on Titan because it is fully three-dimensional and incorporates a detailed microphysics treatment of liquid-methane droplet growth. The dynamical part of the model is adapted to the characteristics of Titan's atmosphere from our anelastic model of moist convection in Jupiter and Saturn^{17,18}. An essential part of the model is the presence of aerosol particles that act as cloud condensation nuclei (CCN). The microphysics scheme used incorporates the processes of condensation, coagulation, evaporation and precipitation¹⁹ as well as coalescence of particles of different sizes falling at different terminal speeds²⁰. For simplicity, no ice processes were considered. The antifreeze effect of nitrogen dissolved into methane results in liquid methane below an altitude of 15 km, and the ice microphysics of methane remain poorly known²¹.

We used the temperature profile derived from Voyager's Infrared Radiometer and Spectrometer (IRIS) data²² and a surface methane molar mixing ratio of 0.05, consistent with the Huygens probe measurements²³. We varied the maximum relative humidity of

methane in the troposphere from 50% to 150%, consistent with previous analysis of the Voyager IRIS data²⁴ and Huygens probe measurements²³ (see Supplementary Information). The lower atmosphere is conditionally unstable (altitudes $z < 20 \text{ km}$) or close to dry instability ($z < 4 \text{ km}$) in the Huygens landing site²⁵. To initiate convection we test different triggering mechanisms at $z = 5\text{--}10 \text{ km}$. In a subsaturated atmosphere we impose either a thermal perturbation, $\Delta T = 0.5\text{--}1 \text{ K}$, or a moderate vertical updraft of 1 m s^{-1} . This vertical motion can be provided by the Hadley cell convergence, resulting in General Circulation Models¹³ or by forced lifting from orographic features.

Clouds form rapidly in supersaturated atmospheres wherever CCNs are available. We considered atmospheres with abundances of particles serving as nucleation sites from $N = 1$ to 50 particles per cm^3 , consistent with the Huygens probe data³. CCN nuclei were either concentrated on the same volume as the thermal-velocity perturbation or available through the entire atmosphere. We studied the cases when CCN are lifted from the surface (and placed in the model at a height of 5 km) or when they settle down from the stratosphere (and placed at a height of 30 km). For simplicity all the perturbed fields in our model have the same volume: a sphere of 2.5 km in radius.

Our microphysics scheme considers a bin distribution of particles with 34 sizes logarithmically distributed between a minimum particle radius of $15 \mu\text{m}$ and a maximum of 4.5 mm, the largest size the drops can attain before breaking by aerodynamic drag¹⁹. The differential sedimentation of particles of different size produces collisions that result in the merging of droplets and particle growth. This process can be incorporated from the expected number of collisions, assuming a given coalescence efficiency from 0 (no coalescence) to 1 (all collisions result in net growth). We do not incorporate particle break-up processes but do not allow particles to grow larger than the upper limiting size either (see Methods).

We run several simulations under different atmospheric conditions (humidity, particle concentration and their spatial distribution) and convective triggering mechanisms (temperature and vertical wind-velocity perturbation) to study the range of conditions that reproduce the observations of the south polar clouds. We also compute the amount of methane precipitation arriving at the surface and its area coverage. Table 1 presents a selection of the most relevant cases able to produce wet-methane convective storms, together with the main properties of the resulting clouds and surface-methane precipitation rates.

Figure 1 presents a comparison between a south polar cloud feature imaged by different instruments and our reference simulation (case S2 in Table 1), in which we consider full efficiency in the coalescence processes. For a 90% relative humidity, a small temperature perturbation of 0.5 K triggers convection, producing methane condensation that heats the parcel through latent heat release and powers vertical motions with updrafts reaching velocities of 18 m s^{-1} . A quick particle growth occurs within the cell with cloud tops reaching 25 km altitude. Figure 2a shows the growth of methane

¹Departamento de Física Aplicada I, ETS Ingenieros, Universidad del País Vasco, Alameda Urquijo s/n, 48013 Bilbao, Spain.

Table 1 | Titan methane storm simulations

Case	Model parameters					Model results			
	CCN location	f	h (%)	N (cm^{-3})	ΔT_0 (K)	H (km)	w (m s^{-1})	Δt rain (h)	P_s (kg m^{-2})
S1	Everywhere	1	90	1	1	13	5	3	12
S2	Everywhere	1	90	10	0.5	25	18	4	110
S3	Everywhere	1	90	10	0	25	17	4	100
S4	Everywhere	1	90	50	0	25	17	4	115
Sd1	10 km	1	125	1	0	12	4	2	35
Sd2	10 km	0	125	1	0	27	16	2.5–8	190
Sd3	30 km	0	125	1	0	30	0	3–10	34
Interacting storms (subsaturated S2 case)									
M2 ($d = 10$ km)	Everywhere	1	90	10	0.5	25	17	3.5	102

S indicates subsaturated atmosphere, Sd indicates saturated atmosphere, and M indicates two interacting cells. CCN location refers to their vertical location, f to the collisional growth efficiency, h is the relative humidity, N the CCN abundance, ΔT_0 the initial thermal perturbation, H the top cloud level, w the maximum updraft velocity, Δt the timescale for development of intense precipitation and P_s the average precipitation at the surface. In case M2, d is the initial distance between the two storms. Additional simulations are in the Supplementary Information.

particles at the 10 km level over the storm lifetime. After two hours, strong precipitation on the same area as the initial perturbation begins (Fig. 2b, c) and after 4 h, most of the condensed methane has rained out with an accumulated precipitation of 110 kg m^{-2} . Later on, cloud droplets continue to fall during several hours from the periphery of the storm. In Fig. 2b and c we also show the time evolution of the precipitation falling to the surface and the area covered by the rain for the cases listed in Table 1.

In general, our results indicate that the onset of moist convection can be produced by moderate perturbations (1 K temperature, 1 m s^{-1} vertical velocity) in any high-relative-humidity atmosphere ($h > 80\%$). If the atmosphere becomes supersaturated, a minimum concentration of CCNs of $\sim 1 \text{ cm}^{-3}$ is required to develop the convective cells. Low CCN concentrations of 0.01 cm^{-3} did not show cloud or convective development. The typical lifetime of a convective cell comprises a growth phase (1–3 h), a mature phase when most rainfall occurs (2–4 h) and a dissipation stage (4–10 h), as shown in Fig. 2b and c.

So far, we have addressed the onset and evolution of a single convective cell. To explain the cloud-covered area observed in Titan's south pole (Fig. 1), the simultaneous development of multiple convective cells is required. Because of the weak convergence at low levels associated to the vertical motions, single convective cumulus can be present at short distances without strongly affecting each other. To test this behaviour, we performed simulations of the interaction between two simultaneous storms separated by a variable distance (case M2 in Table 1). We find that when the distance between cumulus clouds is smaller than 20 km they tend to group together,

attracted by the low pressures associated with the conversion of methane vapour to rain. The approaching storms produce a larger single-cell cloud with a reduced activity (see Supplementary Information). For greater initial separations, the single-cell behaviour is basically preserved and no merger occurs. In such a case, we estimate that about 100 single cells should be sufficient to fit the cloud-covered area of Titan's south pole (Fig. 1). Our results on cloud-top altitude, lifetime and cell separation are consistent with the Cassini and ground-based observations^{1,9}.

The amount of precipitation for the selected cases shown in Table 1 and Fig. 2 is comparable to severe storms on Earth ($\sim 100 \text{ kg m}^{-2}$; ref. 20). The vertical velocities and upper cloud divergences of Titan storms are slightly smaller than the Earth's case, a consequence of the lower gravity and the lower latent heat of condensation of methane compared to water, although partially compensated by the larger abundance of methane and the colder temperatures there. According to current storm models on Earth²⁶ ice processes not included here would generate stronger storms with higher clouds and precipitation. However, the altitude of the highest plumes observed, ~ 40 km at -40° latitude¹¹, cannot be fitted by our models. Even an increase of surface relative humidity from 50% to 95% cannot produce clouds at these heights without assuming an atmosphere significantly less stable than that at the Huygens landing site.

Although the fractional area coverage and the rate of the methane convective storms should be lower on Titan than on Earth¹², our study indicates that violent rain of methane can occur on Titan under the appropriate environmental conditions and could be in the origin of the flow-shaped landscapes observed by Cassini and Huygens^{1,16}.

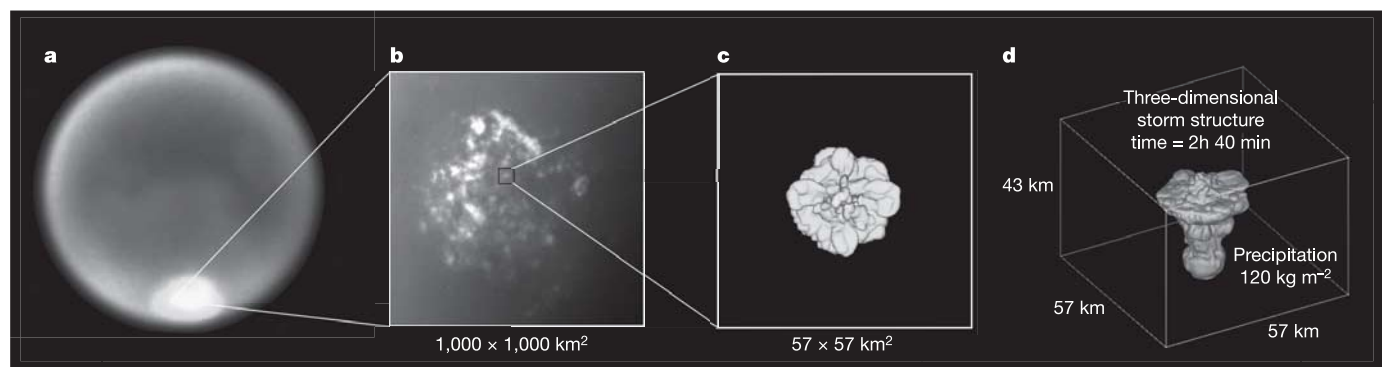


Figure 1 | Observations and numerical simulation of convective clouds in Titan. **a**, Titan image obtained at the Keck Observatory at wavelengths near $2.1 \mu\text{m}$ on 14 November 2003 (ref. 10) showing the intense cloud activity at Titan's south pole. **b**, Cassini high-resolution image obtained on 2 July 2004 (NASA-PDS image PIA06110, ref. 1). **c**, Numerical simulation of a

single methane convective cell (case S2 in Table 1) as viewed from above. **d**, Three-dimensional structure of the cell. The full cloud coverage observed in the Keck and Cassini images can be reproduced by about 100 single cells.

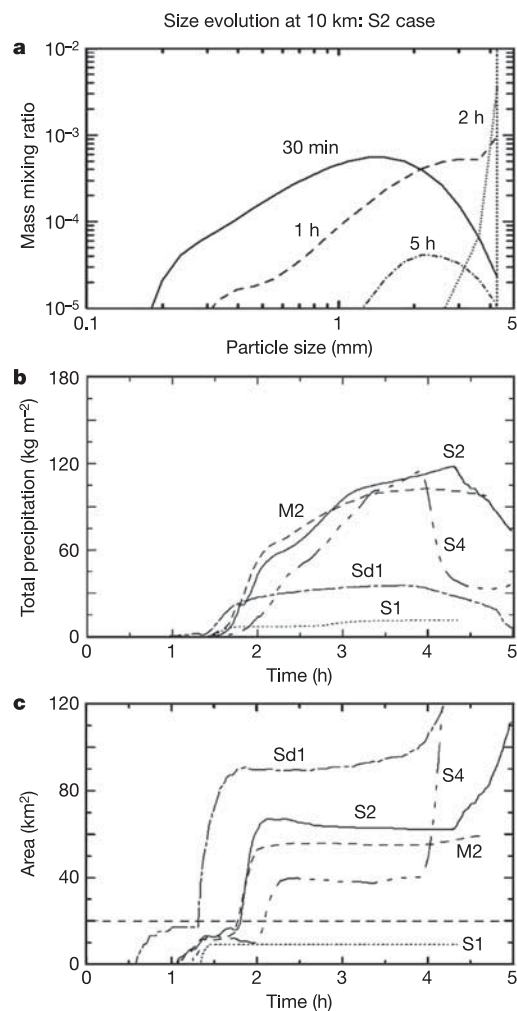


Figure 2 | Particle size evolution and rain in a Titan thunderstorm. **a**, Particle size evolution for the simulated storm shown in Fig. 1. The particles grow quickly to millimetre radius (solid line). After two hours the mass is concentrated in the largest particles (dotted line) and strong precipitation dissipates the cloud, leaving small particles (dashed-dotted line). **b**, Temporal evolution of the precipitation of different simulated storms (see Table 1). Most precipitation occurs after 3–4 h of the onset. **c**, Temporal evolution of the area covered by the precipitation. The horizontal dashed line represents the initial area of the convective disturbance.

We predict that these intense showers should leave a signature on the surface in the form of temporary liquid deposits and surface cooling by ~ 1 K that can be searched by Cassini instruments. Ongoing Cassini and ground-based observations of the cloud activity on Titan will better determine the cloud properties and constrain the model parameters as well as the lifetime of the storms²⁷, helping us to understand the role of convection in the atmospheric methane cycle.

METHODS

Single-storm simulations run in a model domain of $39 \times 39 \text{ km}^2$ and 37 km height with a spatial resolution of 0.5 km. Larger-scale simulations were performed to study the interaction of convective cells in (domain $79 \times 49 \text{ km}^2$ and 49 km height). We used an adaptive time-step of ~ 10 s, satisfying the Courant–Friedrichs–Lewy stability condition²⁸. Boundary conditions were free non-slipping on the lateral sides and a sponge region was introduced in the upper boundary to filter out gravity waves. The lower boundary was set with null velocities and with free non-slipping conditions in the other parameters.

We calculate the evolution of liquid-methane droplets, assuming a logarithmic

bin-size distribution. For each particle size we compute its change by condensation or evaporation. The condensation process is controlled by vapour diffusion under supersaturation conditions and is modulated by the methane surface tension and the release of condensation latent heat. We followed the microphysics scheme described in classic textbooks²⁰, with most of the parameters relevant to methane microphysics taken from ref. 19. Sedimentation of particles was computed as in ref. 19. Each bin precipitates at a different rate, so the particles with larger sizes collide with the smaller particles. For two bins of different size, i and j , the number of collisions per unit volume, N_{cij} , in a given time interval Δt depends on their abundances, sizes and differential sedimentation speed, a process that can be written as:

$$N_{cij} = N_i N_j (\sigma_i + \sigma_j) |V_i - V_j| \Delta t$$

where N_i and N_j are the number density of particles of sizes i and j , σ_i and σ_j are their cross-sections and V_i and V_j are their terminal falling speeds. The efficiency of the collisions in producing larger particles was modulated by a single parameter that controls the degree of particle destruction and growth by differential sedimentation, f ($f = 0$ for no coagulation and 1 for full efficiency of collisional growth). The collisions destroy fN_{cij} particles of sizes i and j and create fN_{cij} particles of mass $m_{ij} = m_i + m_j$. Because the particles evolve to a new size distribution we transform the new particle distribution to the fixed sizes, imposing mass continuity and conservation of the total number of particles.

The simultaneous requirement of considering initial small CCNs and a large bin-size resolution with a reasonable number of sizes (34 bin sizes) imposed on us a lower practical limit of $15 \mu\text{m}$ for the initial cloud droplet population. Initial tests with particles of 1 mm did not show significantly different results.

Received 6 February; accepted 23 May 2006.

- Porco, C. *et al.* An overview of Titan from the Cassini spacecraft. *Nature* **434**, 159–168 (2005).
- Tomasko, M. G. *et al.* Rain, winds and haze during the Huygens probe's descent to Titan's surface. *Nature* **438**, 765–778 (2005).
- Roe, H. G., de Pater, I., Macintosh, B. A. & McKay, C. P. Titan's clouds from Gemini and Keck adaptive optics imaging. *Astrophys. J.* **581**, 1399–1406 (2002).
- Lebreton, J. P. *et al.* An overview of the descent and landing of the Huygens probe on Titan. *Nature* **438**, 758–764 (2005).
- Lorentz, R. D., McKay, C. P. & Lunine, J. I. Photochemically driven collapse of Titan's atmosphere. *Science* **275**, 642–645 (1997).
- Maddox, R. A. Mesoscale convective complexes. *Bull. Am. Meteorol. Soc.* **61**, 1374–1387 (1980).
- Griffith, C. A., Owen, T., Miller, G. A. & Geballe, T. R. Transient clouds in Titan's lower atmosphere. *Nature* **359**, 575–578 (1998).
- Griffith, C. A., Hall, J. L. & Geballe, T. R. Detection of daily clouds on Titan. *Science* **290**, 509–513 (2000).
- Brown, M. E., Bouchez, A. H. & Griffith, C. A. Direct detection of variable tropospheric clouds near Titan's south pole. *Nature* **420**, 795–797 (2002).
- Roe, H. G., Brown, M. E., Schaller, E. L., Bouchez, A. H. & Chadwick, A. Geographic control of Titan's mid-latitude clouds. *Science* **310**, 477–479 (2005).
- Griffith, C. A. *et al.* The evolution of Titan's mid-latitude clouds. *Science* **310**, 474–477 (2005).
- Awai, M. & Lunine, J. I. Moist convective clouds in Titan's atmosphere. *Geophys. Res. Lett.* **21**, 2491–2494 (1994).
- Tokano, T. Meteorological assessment of the surface temperatures on Titan: constraints on the surface type. *Icarus* **173**, 222–242 (2005).
- Lorenz, R. D., Griffith, C. A., Lunine, J. I., McKay, C. P. & Rennò, N. O. Convective plumes and the scarcity of Titan's clouds. *Geophys. Res. Lett.* **32**, L01201, doi:10.1029/2004GL021415 (2005).
- Rannou, P., Montmessin, F., Hourdin, F. & Lebonnois, S. The latitudinal distribution of clouds on Titan. *Science* **311**, 201–205 (2006).
- Elachi, C. *et al.* Cassini radar views the surface of Titan. *Science* **308**, 970–974 (2005).
- Hueso, R. & Sánchez-Lavega, A. A three-dimensional model of moist convection for the giant planets: the Jupiter case. *Icarus* **151**, 257–274 (2001).
- Hueso, R. & Sánchez-Lavega, A. A three-dimensional model of moist convection for the giant planets II: Saturn's water and ammonia moist convective storms. *Icarus* **172**, 255–271 (2004).
- Lorenz, R. The life, death and afterlife of a raindrop on Titan. *Planet. Space Sci.* **41**, 647–655 (1993).
- Houze, R. *Cloud Dynamics* Ch. 3 (Academic Press, San Diego, 1994).
- Thompson, W. R., Zollweg, J. A. & Gabis, D. H. Vapor-liquid equilibrium thermodynamics of $\text{N}_2 + \text{CH}_4$ —Model and Titan applications. *Icarus* **97**, 187–199 (1992).
- Lellouch, E. Titan's atmosphere and hypothesized ocean: A reanalysis of the Voyager 1 radio-occultation and IRIS $7.7 \mu\text{m}$ data. *Icarus* **79**, 328–349 (1989).

23. Niemann, H. B. *et al.* The abundances of constituents of Titan's atmosphere from the GCMS instrument on the Huygens probe. *Nature* **438**, 779–784 (2005).
24. Samuelson, R. E., Nath, N. R. & Borysow, A. Gaseous abundances and methane supersaturation in Titan's troposphere. *Planet. Space Sci.* **45**, 959–980 (1997).
25. Fulchignoni, M. *et al.* *In situ* measurements of the physical characteristics of Titan's environment. *Nature* **438**, 785–791 (2005).
26. Tao, W. K. *et al.* Microphysics, radiation and surface processes in Goddard Cumulus Ensemble (GCE) model. *Meteor. Atmos. Phys.* **82**, 97–137 (2003).
27. Schaller, E. L., Brown, M. E., Roe, H. G. & Bouchez, A. H. A large cloud outburst at Titan's south pole. *Icarus* **182**, 224–229 (2006).
28. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes* 829 (Cambridge Univ. Press, Cambridge, 1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the MCYT Plan Nacional de Astronomía y Astrofísica, Grupos UPV and Fondos FEDER. R.H. was supported by the “Ramón y Cajal” programme. The SGI/IZO-SGIker UPV/EHU is gratefully acknowledged for allocation of computational resources.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.H. (wubhualr@lg.ehu.es) or A.S.-L. (agustin.sanchez@ehu.es).

LETTERS

Methane drizzle on Titan

Tetsuya Tokano¹, Christopher P. McKay², Fritz M. Neubauer¹, Sushil K. Atreya³, Francesca Ferri⁴, Marcello Fulchignoni^{5,6} & Hasso B. Niemann⁷

Saturn's moon Titan shows landscapes with fluvial features¹ suggestive of hydrology based on liquid methane. Recent efforts in understanding Titan's methane hydrological cycle have focused on occasional cloud outbursts near the south pole^{2–4} or cloud streaks at southern mid-latitudes^{5,6} and the mechanisms of their formation. It is not known, however, if the clouds produce rain or if there are also non-convective clouds, as predicted by several models^{7–11}. Here we show that the *in situ* data on the methane concentration and temperature profile in Titan's troposphere point to the presence of layered optically thin stratiform clouds. The data indicate an upper methane ice cloud and a lower, barely visible, liquid methane-nitrogen cloud, with a gap in between. The lower, liquid, cloud produces drizzle that reaches the surface. These non-convective methane clouds are quasi-permanent features supported by the global atmospheric circulation, indicating that methane precipitation occurs wherever there is slow upward motion. This drizzle is a persistent component of Titan's methane hydrological cycle and, by wetting the surface on a global scale, plays an active role in the surface geology of Titan.

The descent of the Huygens probe into Titan's troposphere on 14 January 2005 provided a direct means of determining the condensation state of methane in regions where no clouds have been recognized before. The methane mixing ratio measured by the Huygens Gas Chromatograph Mass Spectrometer (GCMS) is characterized by an essentially uniform methane mixing ratio of $(4.92 \pm 0.25) \times 10^{-2}$ between the surface and 6–8 km altitude, and a monotonic decrease with altitude above this level down to 1.62×10^{-2} at the tropopause^{12,13}. Using the pressure profile measured by the Huygens Atmospheric Structure Instrument (HASI)¹⁴, this yields a column methane abundance of $2,040 \pm 100 \text{ kg m}^{-2}$, equivalent to 2.86 km amagat or ~5 m of liquid methane, which is intermediate between previous values mostly ranging from ~2 km amagat to ~4 km amagat (refs 15, 16), ruling out both excessive supersaturation in the mid-troposphere and a largely subsaturated atmosphere.

We inferred the detailed structure of the implied condensation layer by analysing the relative humidity (RH) of methane (Fig. 1), calculated from the measured methane mixing ratio, temperature and pressure. In the Earth's atmosphere the RH of water rarely exceeds 100%, so regions with an RH of 100% are almost always accompanied by the presence of clouds, fog and/or precipitation. Pure CH₄ freezes at 90.6 K (~2.5 km altitude at Titan), but dissolved nitrogen depresses the freezing point by 10–15 K under Titan conditions—that is, liquid CH₄-N₂ can stably exist up to ~15 km altitude (refs 17, 18). When this effect is taken into consideration, the calculated RH increases almost linearly from ~45% at the surface to 90% at 6 km altitude and then approaches 100% (curve 2 in Fig. 1).

Thus the atmosphere is nearly saturated, implying a cloud above

~8 km altitude that consists of liquid CH₄-N₂ extending at least to the freezing level. The behaviour of this binary CH₄-N₂ mixture below the freezing point is not well determined from experimental data—that is, it is not clear a priori whether the binary mixture can exist as a supercooled liquid, as is often the case in terrestrial clouds at

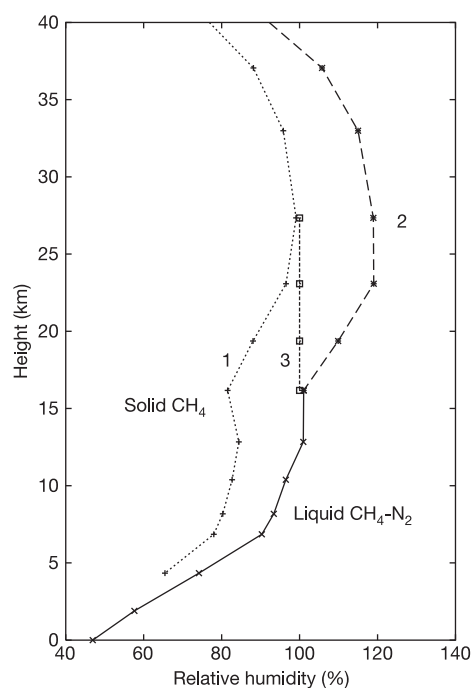


Figure 1 | Vertical profile of the methane relative humidity at the Huygens entry site under different assumptions. The relative humidity (RH) is calculated from the measured methane mixing ratio^{12,13}, temperature¹⁴ and pressure¹⁴. Dashed curve 1 shows the RH with the saturation vapour pressure over pure solid CH₄ (that is, without dissolved N₂). This curve is shown only down to the freezing point of CH₄. The solid curve shows the RH with the saturation vapour pressure over a liquid binary CH₄-N₂ mixture¹⁸, and is shown only up to the freezing level. Dashed curve 2 is the RH with the saturation vapour pressure over a supercooled liquid binary CH₄-N₂ mixture. Dashed curve 3 is a hypothetical RH profile under the assumption that a phase change from liquid to solid and a compositional change from CH₄-N₂ to CH₄ take place. A constant RH of ~100% is possible, even if there is a gap in the cloud. Near the cloud edge turbulent mixing can cause entrainment of drier air from outside the clouds²⁸, so clouds can sometimes exist in subsaturated regions. Therefore, it is not possible to determine the exact altitude of cloud top and bottom solely from the RH profile.

¹Institut für Geophysik und Meteorologie, Universität zu Köln, Albertus-Magnus-Platz, 50923 Köln, Germany. ²NASA Ames Research Center, MS 245-30, Moffett Field, California 94035-1000, USA. ³Department of Atmospheric, Oceanic, and Space Sciences, University of Michigan, 2455 Hayward Street, Ann Arbor, Michigan 48109-2143, USA. ⁴CISAS "G. Colombo", Università di Padova, Via Venezia 15, 35131 Padova, Italy. ⁵LESIA, Observatoire de Paris, 5 Place Jules Janssen, 92195 Meudon, France. ⁶Université Denis Diderot-Paris 7, UFR de Physique, 2 Place Jussieu, 75006 Paris, France. ⁷NASA Goddard Space Flight Center, Code 915, Greenbelt, Maryland 20742, USA.

temperatures below 0 °C. But as upon freezing a substantial fraction of nitrogen should be exsolved from the CH₄ condensate, the saturation vapour pressure would lie somewhere between that over solid pure CH₄ (curve 1 in Fig. 1) and that over liquid CH₄-N₂ (curve 2). Assuming that the CH₄-N₂ condensate freezes near 15 km, we can construct a hypothetical RH profile (curve 3) beginning at 16 km with 100%, and staying constant at this value until it merges with the RH profile for solid CH₄ near 21 km, implying that the N₂ content of the condensate decreases with decreasing temperature. In this case, the RH is virtually uniform from 8 to 30 km, representing one single extensive condensation layer with either an abrupt or successive change in the chemical composition from CH₄-N₂ to CH₄ associated with a phase change above 15 km.

A Titan cloud model that assumes liquid CH₄-N₂ and its freezing¹¹ predicts nearly uniform RH (~100%) extending over several tens of kilometres across the freezing level in the presence of clouds. The resulting upper ice cloud and lower liquid cloud are separated from

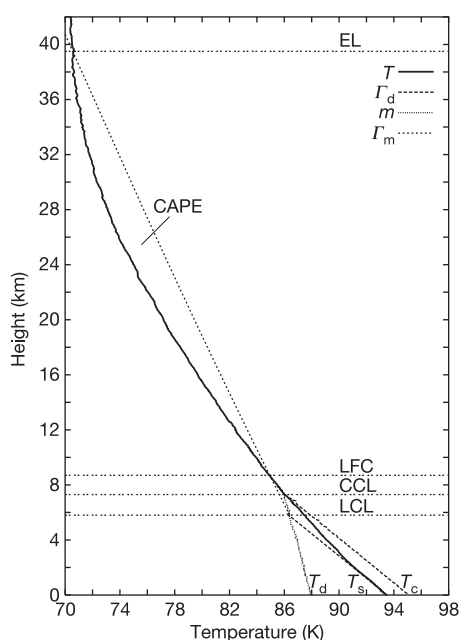


Figure 2 | Thermodynamic diagram at the Huygens entry site. T_s (= 93.6 K) is the surface temperature, T_d (~88 K) is the surface dew point, T_c (~95 K) is the convective temperature (the threshold temperature the surface temperature would have to exceed to enable unforced moist convection for the given methane abundance), LCL (~6 km) is the lifting condensation level, CCL (~9 km) is the convective condensation level, LFC (~7 km) is the level of free convection, EL (~40 km) is the equilibrium level, Γ_d is the dry adiabat, Γ_m is the moist adiabat, T is the measured temperature profile and along the line m the saturation mixing ratio is constant. CAPE (convective available potential energy) is the buoyant energy that is available for rising air parcels that can be converted to kinetic energy of moist convection in the case of successful triggering, and corresponds to the area bounded by the temperature curve and the moist adiabat between LFC and EL. In calculating the dry adiabatic lapse rate, the real gas equation (virial equation)²⁹ is used, considering the atmospheric composition measured by the GCMS¹² and the vertical variation of it; the moist adiabatic lapse rate is calculated with the Redlich-Kwong equation of state³⁰, which includes the effect of the binary CH₄-N₂ mixture. CAPE amounts to ~960 J kg⁻¹, corresponding to weakly unstable convective potential by terrestrial standards, so the chance of severe storms is rather low at the time and place of Huygens landing. The equilibrium level (EL), where ascending air experiences negative buoyancy, is located near 40 km (close to the mean cloud top height of mid-southern-latitude clouds recently observed⁵), indicating that convective clouds, if they develop at all, could indeed extend up to ~40 km. Both clouds are located in a stably stratified environment, so they do not represent inversion clouds of stratocumulus type, but may be classified as stratus-type clouds.

each other by a narrow gap, although this gap does not manifest itself in a local minimum of RH or methane mixing ratio. The lower liquid cloud is mainly a result of melting of the upper ice cloud rather than a product of *in situ* condensation into the liquid phase. With our assumption of compositional and phase change, the upper cloud ice consists of nearly pure solid CH₄, while the lower cloud below 16 km consists of a liquid CH₄-N₂ mixture with a nitrogen concentration of ~20%. The sudden increase in the CH₄ count rate measured by the GCMS near 16 km (refs 12, 13) seems to support this liquid–solid transition; the entry of liquid droplets into the heated inlet of the GCMS can better explain the CH₄ count rate jump than the entry of solid particles. This scenario is consistent with the detection of an optically thin methane haze near 21 km by the Huygens DISR (Descent Imager Spectral Radiometer)¹. The apparent disappearance of the observed methane haze below 20 km indicates a gap in the condensation layer as predicted¹¹, while the top of the methane haze near 21 km suggests a drop in the cloud mass because of decreasing air density with height. However, a subvisible cloud should extend up to the altitude at which the RH drops below 100%, that is, ~30 km. The presence of a cloud gap below 20 km can be understood as an immediate consequence of the phase change: this transitional region is too cold to sustain liquid cloud, but slightly too dry for pure CH₄ to condense.

If we assume that liquid CH₄-N₂ does not freeze but exists as supercooled liquid, the corresponding RH increases above 16 km and attains a maximum of 120% between 23 km and 27 km; that is, there would be substantial supersaturation, as was suggested after the Voyager mission^{19,20}. However, the presence of a vertical gradient in the methane mixing ratio in this altitude region¹³ points to coexistence with liquid condensates. There are two difficulties in supporting this RH profile. First, the detection of a thin methane haze near 21 km (ref. 1) conflicts with the assumption of supercooled liquids in this altitude region. Second, the increase of RH in two separate altitude regions is not predicted by any methane condensation model^{8–11} under any assumption. Furthermore, the presence of supercooled

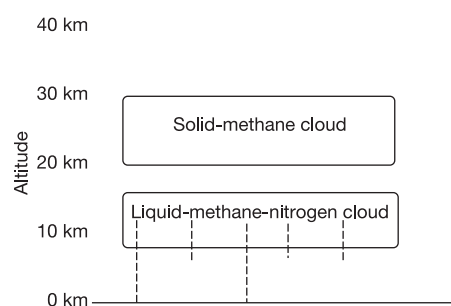


Figure 3 | Diagram of the vertical cloud structure at the Huygens entry site. The upper cloud consists of solid particles, mainly CH₄, and most resembles a terrestrial cirrostratus. The lower cloud consists of a liquid binary CH₄-N₂ mixture and resembles a terrestrial stratus. Either cloud may or may not contain small amounts of C₂H₆ ice as condensation nuclei. However, considering the low downward flux of C₂H₆ ice¹³ and the capability of bare tholins to serve as condensation nuclei for methane¹¹, the role of C₂H₆ may not be as significant as thought before. The upper and lower cloud are separated from each other by a gap because no supercooled droplets exist. CH₄ ice particles falling from the upper cloud passing this gap supply the lower cloud upon melting. Dashed lines indicate rainfall that partly arrives at the surface. Falling drizzle gradually becomes poorer in nitrogen by preferential evaporation of dissolved nitrogen. Except for latitudinal variation in cloud heights, this basic structure may be representative of approximately half of Titan that experiences slow large-scale upwelling. In a colder environment (for example, the winter pole), the lower liquid cloud may extend to higher altitudes, as the freezing level increases¹⁷. If the global circulation pattern changes with season^{9,10}, it will induce a monsoon with a wet season during upwelling and a dry season during downwelling.

liquid $\text{CH}_4\text{-N}_2$ means that neither a phase change nor a compositional change of the cloud would take place, but the kink at the bottom of curve 2 (Fig. 1) near 16 km is suggestive of an abrupt change, contradicting our assumption of no transition.

Similarly, the presence of only pure solid CH_4 is internally inconsistent. The calculated RH is roughly 20% smaller and reaches a plateau (7–16 km) of only $\sim 80\%$, but increases again and eventually reaches saturation (100%) between 25 and 30 km. We might envision a recent occurrence of a large, non-steady condensation event extending from 8 km to 16 km that temporarily suppressed the RH to 80%, but such an event requires or induces a temporal anomaly of several kelvin in the temperature profile that is not observed (Fig. 2). Moreover, the Huygens DISR¹ did not find an optically thick methane cloud around 10 km that would point to such an event. Therefore, we cannot find a physically meaningful mechanism to keep the RH at $\sim 80\%$ for some kilometres below 16 km and to allow an increase of RH to 100%. The only internally consistent scenario is the separate presence of an upper CH_4 ice cloud between about 20 km and 30 km and a lower liquid $\text{CH}_4\text{-N}_2$ cloud between ~ 8 km and 16 km (Fig. 3). Optical detection of the lower liquid cloud by Huygens has not been reported so far, indicating that this cloud may be subvisible. Modelling¹¹ suggests that sub-visible clouds are established by the cloud mass balance if the upward methane flux is slow ($10^{-7} \text{ g cm}^{-2} \text{ s}^{-1}$, corresponding to an eddy diffusivity of $5,000 \text{ cm}^2 \text{ s}^{-1}$) and the droplets are large (0.01–2 mm, “rain without clouds”), and is an indication that this cloud is a steady-state feature, in contrast to many clouds observed elsewhere^{2–6}. The size of cloud droplets varies between less than 0.1 mm (ref. 10) and between 0.1 mm and 1 mm with more particles on the lower side¹¹, so they may be best characterized as drizzle.

The thermodynamic diagram (Fig. 2) reveals that the temperature at the surface was 1–2 K too low to enable free moist convection at the convective condensation level. Forced moist convection at the lifting condensation level near 6 km cannot be definitely ruled out, but we are unable to find a suitable lifting mechanism: orographic forcing is unlikely near the Huygens site, given the flat topography¹ and the weak wind near the surface^{1,21}. Lifting of air along a cold front could not occur owing to the lack of baroclinic instability on Titan⁹; also, the vertical profile of temperature and dew point does not show any anomaly that would point to substantial horizontal advection of heat or moisture. Cryovolcanic activity²² could trigger mid-latitude clouds⁶, but no such features are known in the vicinity of the Huygens site. There is also no evidence of strong updrafts typical of thunderstorms²³.

The driver of this type of condensation is the slow upward transport of methane by large-scale atmospheric circulation^{9,10}. The presence of upward wind in the troposphere at the Huygens site was directly confirmed by HASI²³, and is also predicted by general circulation models (GCMs)^{9,10} for this season.

We note that the vertical temperature profiles at two separate near-equatorial regions retrieved by Voyager 1 were almost identical²⁴, the equator-to-pole temperature contrast was merely 3 K (ref. 20) and the methane abundance determined by remote sensing²⁰ and predicted by GCMs^{9,10} is almost uniform within at least $\pm 30^\circ$ latitude; together, these observations imply that the single measurement by Huygens may be representative of at least 60% of Titan's globe. GCMs^{9,10} predict the widespread presence of such thin stratiform clouds anywhere that slow mean upward motion is found—that is, over about half of Titan's globe at any instant, even if the methane abundance decreases with latitude. Therefore, the clouds encountered by Huygens possibly cover half of Titan.

Previous model predictions^{8–11} can be used to estimate the precipitation rate at the landing site from the measured methane humidity profile. The most important constraint on the precipitation is the uniform methane mixing ratio below 6 km (refs 12, 13), which we

interpret as evidence of slow evaporation of falling rain, as evaporation would produce a vertical gradient in the methane mixing ratio^{8,9,11}. Hence the rain from the liquid cloud must eventually pass through the lower troposphere and reach the surface to achieve mass balance, because otherwise the cloud would grow sufficiently to be detected optically. We estimate the surface precipitation rate at the Huygens landing site as $\sim 50 \text{ mm yr}^{-1}$ (see Supplementary Information).

On Earth, regions with this annual precipitation rate are considered deserts, but the rainfall on Titan is inferred to be more persistent than in terrestrial deserts, given the stratiform character of the clouds. With the measured column methane abundance and this precipitation rate, the atmospheric residence time is estimated as 80 years, which is 5–6 orders of magnitude shorter than the photochemical lifetime of Titan's atmospheric methane. Non-zero surface precipitation means that the hydrological cycle of atmospheric methane regularly involves the surface, contrary to some previous expectations. Although the erosive potential may be quite limited with such a low precipitation rate²⁵, it would be at least sufficient to wet the surface material, and may explain the generally wet character of the surface material²⁶. Atmospheric methane can probably be supplied by (cryovolcanic) outgassing^{13,22,27}, but the atmospheric circulation is required to globally distribute methane and precipitation^{9,10} and thus produce global wetting of the surface. Large-scale stratiform precipitation—drizzle—may constitute a more persistent component of Titan's whole methane cycle¹³ than the optically thick but sporadic clouds.

Received 29 March; accepted 6 June 2006.

- Tomasko, M. G. *et al.* Rain, winds and haze on Titan during the Huygens probe's descent to Titan's surface. *Nature* **438**, 765–778 (2005).
- Brown, M. E., Bouchez, A. H. & Griffith, C. A. Direct detection of variable tropospheric clouds near Titan's south pole. *Nature* **420**, 795–797 (2002).
- Porco, C. C. *et al.* Imaging of Titan from the Cassini spacecraft. *Nature* **434**, 159–168 (2005).
- Schaller, E. L., Brown, M. E., Roe, H. G. & Bouchez, A. H. A large cloud outburst at Titan's south pole. *Icarus* **182**, 224–229 (2006).
- Griffith, C. A. *et al.* The evolution of Titan's mid-latitude clouds. *Science* **310**, 474–477 (2005).
- Roe, H. G., Brown, M. E., Schaller, E. L., Bouchez, A. H. & Trujillo, C. A. Geographic control of Titan's mid-latitude clouds. *Science* **310**, 477–479 (2005).
- Toon, O. B., McKay, C. P., Courtin, R. & Ackerman, T. P. Methane rain on Titan. *Icarus* **75**, 255–284 (1988).
- Samuelson, R. E. & Mayo, L. A. Steady-state model for methane condensation in Titan's troposphere. *Planet. Space Sci.* **45**, 949–958 (1997).
- Tokano, T., Neubauer, F. M., Laube, M. & McKay, C. P. Three-dimensional modeling of the tropospheric methane cycle on Titan. *Icarus* **153**, 130–147 (2001).
- Rannou, P., Montmessin, F., Hourdin, F. & Lebonnois, S. The latitudinal distribution of clouds on Titan. *Science* **311**, 201–205 (2006).
- Barth, E. L. & Toon, O. B. Methane, ethane, and mixed clouds in Titan's atmosphere: properties derived from microphysical modeling. *Icarus* **182**, 230–250 (2006).
- Niemann, H. B. *et al.* The abundances of constituents of Titan's atmosphere from the GCMS instrument on the Huygens probe. *Nature* **438**, 779–784 (2005).
- Atreya, S. K. *et al.* Titan's methane cycle. *Planet. Space Sci.* (in the press).
- Fulchignoni, M. *et al.* *In situ* measurements of the physical characteristics of Titan's environment. *Nature* **438**, 785–791 (2005).
- Lemmon, M. T., Smith, P. H. & Lorenz, R. D. Methane abundance on Titan, measured by the Space Telescope Imaging Spectrograph. *Icarus* **160**, 375–385 (2002).
- Penteado, P. F., Griffith, C. A., Greathouse, T. K. & de Bergh, C. Measurements of CH_3D and CH_4 in Titan from infrared spectroscopy. *Astrophys. J.* **629**, L53–L56 (2005).
- Lorenz, R. D. & Lunine, J. I. Titan's snowline. *Icarus* **158**, 557–559 (2002).
- Thompson, W. R., Zollweg, J. A. & Gabis, D. H. Vapor-liquid equilibrium thermodynamics of $\text{N}_2 + \text{CH}_4$: Model and Titan applications. *Icarus* **97**, 187–199 (1992).
- Courtin, R., Gautier, D. & McKay, C. P. Titan's thermal emission spectrum. *Icarus* **114**, 144–162 (1995).
- Samuelson, R. E., Nath, N. R. & Borysow, A. Gaseous abundances and methane supersaturation in Titan's troposphere. *Planet. Space Sci.* **45**, 959–980 (1997).
- Bird, M. K. *et al.* The vertical profile of winds on Titan. *Nature* **438**, 800–802 (2005).

22. Sotin, C. *et al.* Release of volatiles from a possible cryovolcano from near-infrared imaging of Titan. *Nature* **435**, 786–789 (2005).
23. Mäkinen, J. T. T. *et al.* Vertical atmospheric flow on Titan as measured by the HASI instrument. *Geophys. Res. Lett.* (submitted).
24. Lindal, G. F. *et al.* The atmosphere of Titan—an analysis of the Voyager 1 radio occultation measurements. *Icarus* **53**, 348–363 (1983).
25. Lorenz, R. D. & Lunine, J. I. Erosion on Titan: past and present. *Icarus* **122**, 79–91 (1996).
26. Zarnecki, J. C. *et al.* A soft solid surface on Titan as revealed by the Huygens Surface Science Package. *Nature* **438**, 792–795 (2005).
27. Tobie, G., Lunine, J. I. & Sotin, C. Episodic outgassing as the origin of atmospheric methane on Titan. *Nature* **440**, 61–64 (2006).
28. Pruppacher, H. R. & Klett, J. D. *Microphysics of Clouds and Precipitation* (Kluwer, Dordrecht, 1997).
29. Mäkinen, T. Processing the HASI measurements. *Adv. Space Res.* **17**, 217–222 (1996).
30. McKay, C. P., Martin, S. C., Griffith, C. A. & Keller, R. M. Temperature lapse rate and methane in Titan's troposphere. *Icarus* **129**, 498–505 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements T.T. and F.M.N. were supported by the DLR in the HASI project. T.T. received a grant from the DFG. S.K.A. and C.P.M. acknowledge support from NASA's Planetary Atmospheres Program and the Cassini-Huygens Project. M.F. and F.F. acknowledge support from the Italian Space Agency (ASI) for the HASI experiment on board the ESA Huygens probe.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.T. (tokano@geo.uni-koeln.de).

LETTERS

Atom-by-atom substitution of Mn in GaAs and visualization of their hole-mediated interactions

Dale Kitchen^{1,2}, Anthony Richardella^{1,2}, Jian-Ming Tang³, Michael E. Flatté³ & Ali Yazdani¹

The discovery of ferromagnetism in Mn-doped GaAs¹ has ignited interest in the development of semiconductor technologies based on electron spin and has led to several proof-of-concept spintronic devices^{2–4}. A major hurdle for realistic applications of $\text{Ga}_{1-x}\text{Mn}_x\text{As}$, or other dilute magnetic semiconductors, remains that their ferromagnetic transition temperature is below room temperature. Enhancing ferromagnetism in semiconductors requires us to understand the mechanisms for interaction between magnetic dopants, such as Mn, and identify the circumstances in which ferromagnetic interactions are maximized⁵. Here we describe an atom-by-atom substitution technique using a scanning tunnelling microscope (STM) and apply it to perform a controlled study at the atomic scale of the interactions between isolated Mn acceptors, which are mediated by holes in GaAs. High-resolution STM measurements are used to visualize the GaAs electronic states that participate in the Mn–Mn interaction and to quantify the interaction strengths as a function of relative position and orientation. Our experimental findings, which can be explained using tight-binding model calculations, reveal a strong dependence of ferromagnetic interaction on crystallographic orientation. This anisotropic interaction can potentially be exploited by growing oriented $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ structures to enhance the ferromagnetic transition temperature beyond that achieved in randomly doped samples.

In $\text{Ga}_{1-x}\text{Mn}_x\text{As}$, substituted Mn atoms at Ga sites act as acceptors donating holes that are believed to be responsible for mediating ferromagnetic interactions between Mn *d*-orbital core spins. Mean-field theories based on this scenario have captured the increase of the ferromagnetic transition temperature, T_c , with doping and other macroscopic properties of $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ (refs 5, 6). Experimental efforts have also shown that improved material quality, achieved through various methods including annealing of random alloy samples, can significantly increase T_c even above 150 K (refs 7–10). However, we lack an accurate microscopic picture that will identify the precise nature of hole states involved and provide clues on how to enhance ferromagnetism in this compound. Previous STM experiments have mapped the shape of single acceptor states for Mn in subsurface sites¹¹, but have been unable to observe interactions between them directly¹². First-principles calculations^{13,14} have provided insight, but their predictions for the structure of Mn–Mn interactions could only be tested statistically in macroscopic materials that also contained large numbers of defects, interstitials and other complicating features.

We performed our experiments using a home-built cryogenic STM that operates at 4 K in ultrahigh vacuum. We used wafers of GaAs doped with 10^{18} – 10^{19} Zn atoms cm^{-3} , which are cleaved *in situ* to expose a (110) surface. The cleaved samples show STM topography and spectroscopy that are characteristic of a degenerately doped p-type GaAs sample, with the Fermi level very close to the valence

band edge of the GaAs¹⁵. Although the (110) surface undergoes a small reconstruction, there are no surface states at energies within the GaAs gap to complicate our studies¹⁶. To substitute Mn atoms for Ga, we first deposited a small concentration of Mn atoms (0.1–2% monolayer) from an *in situ* source onto the cold sample surface. The deposited Mn atoms are weakly adsorbed on the surface, appearing in valence-band images (tip-sample bias $V < 0$) of a GaAs (110) surface as small protrusions (Fig. 1a and b) and can be manipulated using the STM tip.

We discovered that application of a voltage pulse in the STM junction, which injects energetic electrons onto a Mn adsorbate's site, results in the substitution of one Mn for one Ga atom in the surface layer. This substitution process requires us to inject electrons with energies of 0.7 eV or higher. Varying the precise placement of the tip near the adsorbate resulted in either its substitution or its lateral motion on the surface. Figure 1c and d, after tip-voltage pulses, reveal the presence of the substituted Mn by showing its influence on neighbouring As atoms and the ejection of the Ga atom it has replaced. The precise physical process for the STM-assisted substitution is still under investigation; nonetheless, model calculations show that the substituted configuration of Mn has a lower energy than the adsorbate configuration on the GaAs (110) surface¹⁷. We found that the STM-assisted incorporation and motion of Mn adsorbates can be used to substitute Mn atoms successfully at precise

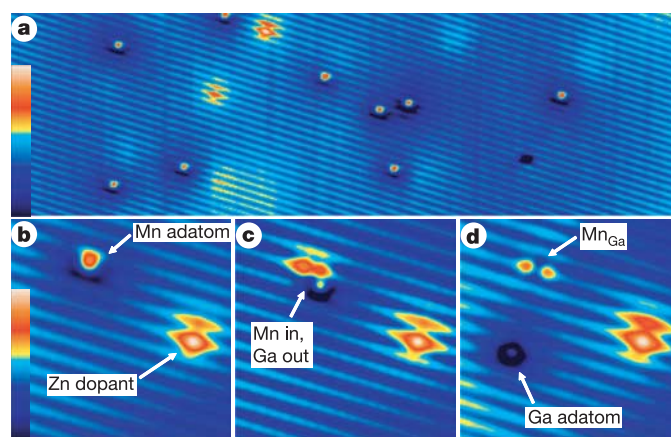


Figure 1 | Single-atom substitution of one Mn for one Ga atom. **a**, Large topograph of the occupied states ($500 \times 150 \text{ Å}^2$; -2 V ; $0\text{--}1.5 \text{ Å}$) showing many Mn adsorbates and a few subsurface Zn dopants. **b–d**, High-resolution topographs (75 Å^2 ; -2 V ; $0\text{--}1.3 \text{ Å}$). **b**, Mn adsorbate and a Zn acceptor visible on the As sublattice. **c**, STM-assisted incorporation of the Mn into a Ga site forcing the substituted Ga atom to the surface. **d**, Tip-pulses move the Ga adatom from the incorporation site, isolating a Mn acceptor (Mn_{Ga}).

¹Department of Physics, Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08544, USA. ²Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. ³Optical Science and Technology Center and Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa 52242, USA.

Ga sites and to remove the ejected Ga atoms from the substitution area. All the experimental results reported here were repeated for Mn at different locations on the surface to ensure that interaction with native Zn acceptors in the substrate did not alter our experimental findings.

The modifications of the local density of electronic states (LDOS) of GaAs due to Mn substitution significantly affect the nature of their interactions within the GaAs host. We probed these modifications by combining real space imaging together with spatially resolved STM spectroscopy, performed by measuring the differential conductance, dI/dV versus V (where I is current and V is voltage), using standard lock-in techniques. The spectroscopic measurements performed in the vicinity of an isolated substituted Mn are shown in Fig. 2a. The modification of the valence-band electronic states ($V < 0$ in the spectra), as indicated by an arrow in the tunnelling spectra of Fig. 2a, is strongest on the As neighbours nearest the Mn. The spatial extent of the Mn-induced modifications of the valence-band states is weak beyond the nearest As neighbours and is spatially anisotropic, as shown in Fig. 2b. The Mn has a greater influence, however, owing to the binding of an acceptor state to the Mn site producing a strong resonance in the tunnelling spectra inside the GaAs gap (Fig. 2a). The dominance of the Mn acceptor state in the LDOS lends itself to direct mapping of the acceptor-state wavefunction in the unoccupied state images. Such images, as shown in Fig. 2c, demonstrate that the Mn acceptor has an anisotropic star-shaped spatial structure that is distributed over more than a $20 \text{ \AA}^2$ area of the surface¹⁸. The large energy width ($>150 \text{ meV}$) of the conductance peak shows that the Mn acceptor state has a large overlap with the continuum states, such

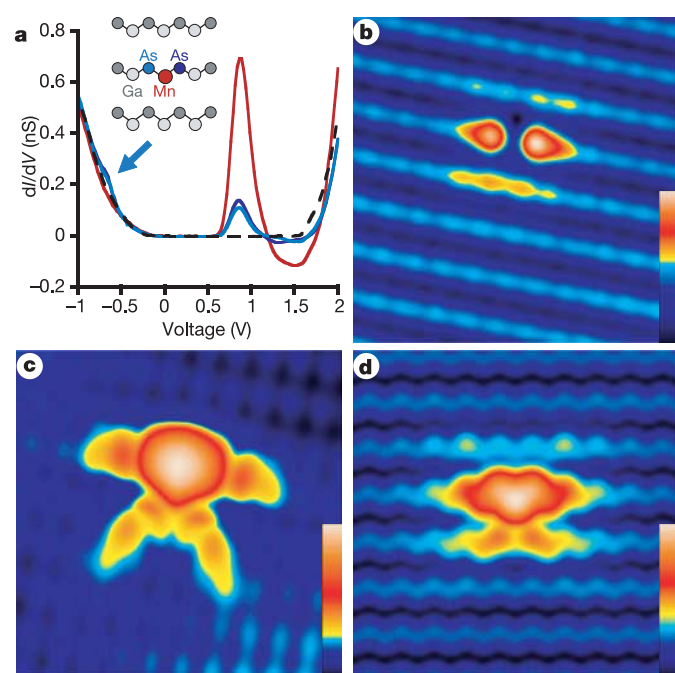


Figure 2 | High-resolution measurements of a single Mn acceptor (Mn_{Ga}). **a**, Spatially resolved dI/dV measurements near an Mn acceptor. The As neighbours show enhancements deep in the valence band, highlighted by the blue arrow. The large in-gap resonance over the Mn is its acceptor level. **b**, $40 \text{ \AA}^2$ topograph of the occupied states (-1.5 V ; $0-0.4 \text{ \AA}$) showing enhancements concentrated on the neighbouring As. **c**, $40 \text{ \AA}^2$ topograph of the unoccupied states ($+1.55 \text{ V}$; $0-4 \text{ \AA}$) revealing the anisotropic shape of the Mn acceptor. **d**, Tight-binding model of a Mn acceptor in bulk GaAs with the Mn spin oriented in-plane along $[001]$. The image shows a $40 \text{ \AA}^2$ area of the (110) plane containing the Mn. The log of the integrated LDOS from the Fermi level to 1.55 eV is displayed with a spatial broadening factor of $1.7 \text{ \AA}$. The colours in the scales have been chosen to enhance the view of the Mn influence on GaAs.

as those due to Zn acceptors, in our substrates, eliminating the possibility of charging effects when tunnelling through this state.

The measured modifications of the LDOS, induced by Mn in the (110) surface layer of GaAs, are consistent with tight-binding calculations of the electronic states in bulk GaAs near Mn acceptors (see Supplementary Information)¹⁹. Figure 2d shows a simulated STM image of the acceptor state based on the bulk tight-binding calculations that model the experimental situation by showing the calculated electronic states near a Mn dopant in the bulk at the (110) layer containing the Mn. These calculations have essentially one free parameter, the p - d hybridization induced by the Mn site, which is adjusted to match the measured bulk acceptor level energy. The spatial structure of the calculated wavefunction for the bulk acceptor shows a similar anisotropic shape and extent as seen in the STM experiment for Mn at the surface. The tight-binding calculations also predict resonances deep within the valence band¹⁹, which account for the valence-band modification imaged in Fig. 2b. These resonance states have been anticipated¹⁹ to play a part in interpretation of photoemission²⁰ and infrared optical conductivity^{21,22} measurements of $\text{Ga}_{1-x}\text{Mn}_x\text{As}$.

More detailed comparisons of our measurements with the bulk calculations and with other STM measurements of subsurface Mn¹¹ provide ways to assess the significance of the surface in our findings. Simulating the impact of the surface on the acceptor state in the bulk calculation by adjusting the potential energy of one of the As bonded to the Mn, we find that the acceptor state shifts deeper into the gap. Our experimentally measured acceptor state is indeed found deeper in the gap (850 meV), as compared to the measured value for the bulk (113 meV), although a significant part of the observed shift is due to tip-induced band-bending typical of tunnelling between a metal and a semiconductor^{15,16}. More importantly, contrasting our measurements with recent STM studies of $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ samples grown by molecular-beam epitaxy¹¹, the subsurface Mn acceptors show states with similar anisotropic symmetry and spatial extent to those reported here. In the studies of subsurface Mn, tunnelling through the layers of GaAs broadens some of the finer features of the acceptor state compared to that reported here—a behaviour captured by our tight-binding model (see Supplementary Information). Overall, comparison of our data to bulk calculations and to previous STM measurements of Mn in subsurface sites shows that surface effects do not significantly alter the acceptor state's characteristics. This comparison motivates the study of the interaction between STM-substituted Mn atoms as a model situation for probing the nature of Mn–Mn interactions mediated through the GaAs host.

Our atom-by-atom substitution technique provides a powerful method of implanting Mn acceptors at precise relative separations and orientations in the GaAs surface and to examine their interaction in pairs. Figure 3 details measurements of an STM-substituted Mn pair, separated by $8 \text{ \AA}$ along a $\langle 110 \rangle$ crystallographic direction. Figure 3a and b shows changes in topography induced in the GaAs valence band and in-gap states by the Mn pair. Direct evidence for interaction within the pair can be observed in the tunnelling spectra of Fig. 3c, which reveals splitting of the acceptor state into two resonances. Figure 3d and e shows results of conductance mapping, which probes the spatial characteristic of electronic states by visualizing changes in dI/dV at specific energies, confirming that the split states observed in the spectra (Fig. 3c) indeed have bonding/anti-bonding characteristics. In this situation, the observed bonding state occurs at higher energies because of the hole-like nature of the states involved, for which the continuum is at the top of the valence band. The observed splitting in Fig. 3 also indicates that the spin states associated with the Mn's $3d$ orbitals are ferromagnetically aligned. Because the formation of bonding and anti-bonding states requires electronic states that are degenerate in both energy and spin alignment, anti-ferromagnetically aligned Mn pairs would have anti-aligned acceptor states that would not show any splitting²³. Within a doping-dependent superexchange model of magnetism (for example, ref. 24),

this state splitting is the dominant contribution to the Mn–Mn spin interaction.

Constructing Mn pairs using our atom-by-atom substitution technique and measuring the energy splitting of their acceptor states provides a method of mapping the relative strength of Mn–Mn interactions as a function of distance and crystallographic orientation in GaAs. A subset of the Mn pairs we have examined is shown in Fig. 4a–d and Fig. 3b, along with measurements of the splitting energy of the acceptor levels for Mn pair configurations up to the six nearest neighbours in Fig. 4e. Our key observation is that the Mn–Mn interaction decays rapidly with increasing separation between the Mn acceptors and is highly anisotropic. The data in Fig. 4e clearly show that pairs constructed along $\langle 110 \rangle$ crystallographic directions have a much stronger splitting than those along the $\langle 100 \rangle$ or $\langle 111 \rangle$ directions. Our ability to probe the energy splitting of the pairs is limited by the energy width of a single Mn acceptor state. Consequently, based on experimental measurements alone, we cannot rule out the possibility that pairs, such as the one separated by 5.65 Å along a $\langle 100 \rangle$ direction, that do not show splitting are not anti-ferromagnetically aligned (see Supplementary Information). Nonetheless, the experimental measurements of the relative strength of the Mn–Mn interactions and their anisotropic character can be used as a test bed for microscopic models of ferromagnetic interactions in GaAs.

As a starting point for understanding our experimental findings for Mn pairs, we turn again to bulk tight-binding calculations, which successfully accounted for the Mn-induced changes in the LDOS of GaAs (Fig. 2d). We calculate splitting of the acceptor states for two bulk Mn atoms at the five nearest-neighbour pairs probed in the STM experiments, assuming that their core spins are ferromagnetically aligned and point along a $\langle 100 \rangle$ direction, the bulk easy axis. As shown in Fig. 4e, the results of these tight-binding calculations agree remarkably well in the overall trend and the observed anisotropy of the experimental data. The discrepancy in the actual values of acceptor-level energy splitting is probably a result of the non-trivial shift of these levels due to tip-induced band-bending in the experiment. The tight-binding model can also be used to estimate the energy gained for ferromagnetic alignment and its connection to the acceptor-state energy splitting (see Supplementary Information).

Looking beyond the tight-binding model, first-principles *ab initio* calculations also predict the same anisotropy in ferromagnetic interactions²⁴ that we find in our STM measurements of the state splittings for the first several nearest-neighbour pairs. In fact, various theoretical models that include detailed band structure of GaAs or the effects of spin-orbit coupling predict anisotropic Mn–Mn interactions of various degrees^{25,26}. Detailed examination of our experimental data for a single Mn shows that the origin of the anisotropic splittings is most probably the shape of the acceptor wavefunction, which favours resonant interaction along the $\langle 110 \rangle$ crystallographic directions. However, valence-band modification of the electronic states due to Mn has a similar anisotropy, which could also mediate anisotropic interactions with a similar symmetry. In addition, spin-orbit interactions also produce crystalline magnetic anisotropy in GaMnAs. Our tight-binding model, which can be used to estimate this anisotropy for Mn pairs, shows that their spin alignment favours the $\langle 100 \rangle$ direction at the surface (see Supplementary Information).

Besides providing experimental results for testing models of Mn–Mn interactions in GaAs, our experiments provide microscopic information that potentially can be used to enhance ferromagnetism in $\text{Ga}_{1-x}\text{Mn}_x\text{As}$. The significance of the experiments is perhaps best realized by considering the separations between Mn acceptors in randomly doped $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ at concentrations required for the highest reported ferromagnetic temperatures^{7–10}. At 5% doping, 98% of the Mn dopants have one or more neighbour at the sixth nearest-neighbour or closer. The experiments reported here suggest that randomly distributed Mn doping fails to take advantage of the large strength of the interaction along the $\langle 110 \rangle$ direction. MBE-grown heterostructures could be optimized to increase the concentration of Mn dopants with $\langle 110 \rangle$ neighbours, and potentially enhance ferromagnetic Mn–Mn coupling. In addition, the possibility that some of the closely spaced pairs, such as the one separated by 5.65 Å along a $\langle 100 \rangle$ direction, could be antiferromagnetically aligned raises the question of whether spin frustration²⁵ could be hampering efforts to increase the ferromagnetic transition temperature in the randomly doped system.

The atom-by-atom substitution technique we have demonstrated

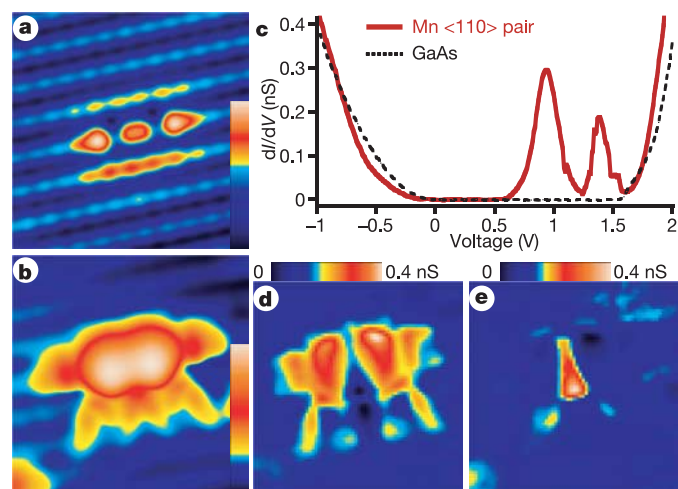


Figure 3 | Mn pair substituted in Ga sites spaced 8 Å apart in a $\langle 110 \rangle$ orientation. **a**, 40 Å² topograph of the valence-band states (−1.5 V; 0–0.4 Å). **b**, 40 Å² topograph of the in-gap states showing the bound states of Mn acceptors (+1.1 V; 0–4 Å). **c**, dI/dV measurements near the Mn pair, revealing two resonance levels in the gap. **d**, Differential conductance energy map at +1.35 V displaying the bonding nature of the higher energy resonance. **e**, Energy map at +0.91 V displaying the anti-bonding nature of the lower energy peak.

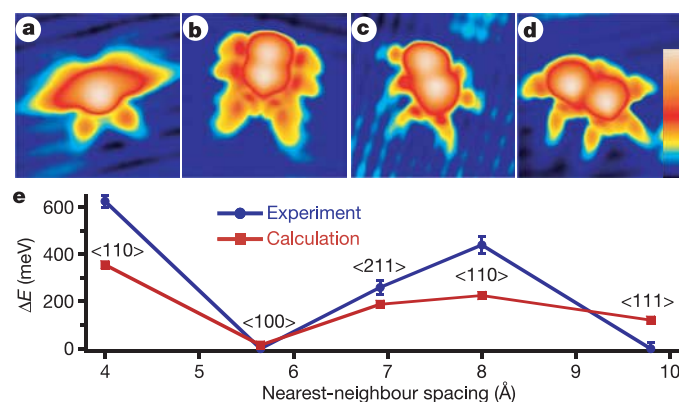


Figure 4 | Topographs (40 Å², +1.5 V) of several Mn pairs and the acceptor-level splitting energy of these pairs. **a**, Nearest-neighbour Mn acceptor pair (4 Å spacing, $\langle 110 \rangle$ orientation). **b**, Second-nearest-neighbour pair (5.65 Å spacing, $\langle 100 \rangle$ orientation). **c**, Third-nearest pair (6.92 Å spacing, $\langle 211 \rangle$ orientation). **d**, Sixth-nearest-neighbour pair (9.79 Å spacing, $\langle 111 \rangle$ orientation). Figure 3b shows the fourth-nearest-neighbour pair (8 Å spacing, $\langle 110 \rangle$ orientation). The fifth-nearest-neighbour is not accessible in a single (110) plane. **e**, Comparative plot of the Mn acceptor-level splitting for Mn–Mn pair interactions. Theory and experiment both show that Mn–Mn interactions are highly anisotropic, favouring $\langle 110 \rangle$ orientations. Error bars correspond to the standard deviation of many pairs measured on p-type GaAs (110) surfaces. ΔE , splitting energy of the acceptor levels.

here can also be an important tool in a number of new directions. A direct extension would be to use our technique to substitute other dopants in various semiconducting surfaces, as a method to search for strongly interacting ferromagnetic donors or acceptors. Such an effort can be guided by theoretical efforts similar to those described here. Single dopant substitution also provides a controlled method to create single spin quantum bits in semiconductors²⁷. The tight-binding calculations show the exciting possibility that the spin-orbit coupling can dictate the shape of a single acceptor wavefunction, hence providing a method for obtaining spin information from measurements of LDOS distribution²⁸.

Received 3 April; accepted 6 June 2006.

- Ohno, H. *et al.* (Ga,Mn)As: A new diluted magnetic semiconductor based on GaAs. *Appl. Phys. Lett.* **69**, 363–365 (1996).
- Ohno, H. *et al.* Electric-field control of ferromagnetism. *Nature* **408**, 944–946 (2000).
- Ohno, Y. *et al.* Electrical spin injection in a ferromagnetic semiconductor device heterostructure. *Nature* **402**, 790–792 (1999).
- Eid, K. F. *et al.* Exchange biasing of the ferromagnetic semiconductor $\text{Ga}_{1-x}\text{Mn}_x\text{As}$. *Appl. Phys. Lett.* **85**, 1556–1558 (2004).
- Macdonald, A. H., Schiffer, P. & Samarth, N. Ferromagnetic semiconductors: moving beyond (Ga,Mn)As. *Nature Mater.* **4**, 195–202 (2005).
- Dietl, T., Ohno, H., Matsukura, F., Cibert, J. & Ferrand, D. Zener model description of ferromagnetism in zinc-blende magnetic semiconductors. *Science* **287**, 1019–1022 (2000).
- Edmonds, K. *et al.* High-Curie-temperature $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ obtained by resistance-monitored annealing. *Appl. Phys. Lett.* **81**, 4991–4993 (2002).
- Ku, K. C. *et al.* Highly enhanced Curie temperature in low-temperature annealed [Ga,Mn]As epilayers. *Appl. Phys. Lett.* **82**, 2302–2304 (2003).
- Chiba, D., Takamura, K., Matsukura, F. & Ohno, H. Effect of low-temperature annealing on (Ga,Mn)As trilayer structures. *Appl. Phys. Lett.* **82**, 3020–3022 (2003).
- Edmonds, K. W. *et al.* Mn interstitial diffusion in (Ga, Mn)As. *Phys. Rev. Lett.* **92**, 037201 (2004).
- Yakunin, A. M. *et al.* Spatial structure of an individual Mn acceptor in GaAs. *Phys. Rev. Lett.* **92**, 216806 (2004).
- Yakunin, A. M. *et al.* Spatial structure of Mn-Mn acceptor pairs in GaAs. *Phys. Rev. Lett.* **95**, 256402 (2005).
- Sanvito, S., Theurich, G. & Hill, N. A. Density functional calculations for III–V diluted ferromagnetic semiconductors: a review. *J. Supercond.* **15**, 85–104 (2002).
- Mahadevan, P. & Zunger, A. First-principles investigation of the assumptions underlying model-Hamiltonian approaches to ferromagnetism of 3d impurities in III–V semiconductors. *Phys. Rev. B* **69**, 115211 (2004).
- Feenstra, R. M. & Stroscio, J. A. Atom-selective imaging of the GaAs(110) surface. *Phys. Rev. Lett.* **58**, 1192–1195 (1987).
- Feenstra, R. M. & Stroscio, J. A. Tunneling spectroscopy of the GaAs(110) surface. *J. Vac. Sci. Technol. B* **5**, 923–929 (1987).
- Fu, H., Ye, L., Zhang, K. & Xie, X. Chemisorption of Mn on a GaAs(110) surface. *Surf. Sci.* **341**, 273–281 (1995).
- Kitchen, D., Richardella, A. & Yazdani, A. Spatial structure of a single Mn impurity state on GaAs (110) Surface. *J. Supercond.* **18**, 23–28 (2005).
- Tang, J.-M. & Flatté, M. E. Multiband tight-binding model of local magnetism in $\text{Ga}_{1-x}\text{Mn}_x\text{As}$. *Phys. Rev. Lett.* **92**, 047201 (2004).
- Okabayashi, J. *et al.* Angle-resolved photoemission study of $\text{Ga}_{1-x}\text{Mn}_x\text{As}$. *Phys. Rev. B* **64**, 125304 (2001).
- Singley, E. J., Kawakami, R., Awschalom, D. D. & Basov, D. N. Infrared probe of itinerant ferromagnetism in $\text{Ga}_{1-x}\text{Mn}_x\text{As}$. *Phys. Rev. Lett.* **89**, 097203 (2002).
- Burch, K. S. *et al.* Impurity band conduction in a high temperature ferromagnetic semiconductor. Preprint at (<http://arXiv.org/cond-mat/0603851>) (2006).
- Flatté, M. E. & Reynolds, D. E. Local spectrum of a superconductor as a probe of interactions between magnetic impurities. *Phys. Rev. B* **61**, 14810–14813 (2000).
- Mahadevan, P., Zunger, A. & Sarma, D. D. Unusual directional dependence of exchange energies in GaAs diluted with Mn: Is the RKKY description relevant? *Phys. Rev. Lett.* **93**, 177201 (2004).
- Zaránd, G. & Jankó, B. $\text{Ga}_{12x}\text{Mn}_x\text{As}$: A frustrated ferromagnet. *Phys. Rev. Lett.* **89**, 047201 (2002).
- Timm, C. & MacDonald, A. H. Anisotropic exchange interactions in III–V diluted magnetic semiconductors. *Phys. Rev. B* **71**, 155206 (2005).
- Kane, B. M. A silicon-based nuclear spin quantum computer. *Nature* **393**, 133–137 (1998).
- Tang, J.-M. & Flatté, M. E. Spin-orientation-dependent spatial structure of a magnetic acceptor state in a zinc-blende semiconductor. *Phys. Rev. B* **72**, 161315(R) (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the US ARO MURI and the US NSF.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.Y. (yazdani@princeton.edu).

LETTERS

Widespread active detachment faulting and core complex formation near 13° N on the Mid-Atlantic Ridge

Deborah K. Smith¹, Johnson R. Cann² & Javier Escartín³

Oceanic core complexes are massifs in which lower-crustal and upper-mantle rocks are exposed at the sea floor^{1–3}. They form at mid-ocean ridges through slip on detachment faults rooted below the spreading axis^{2–6}. To date, most studies of core complexes have been based on isolated inactive massifs that have spread away from ridge axes. Here we present a survey of the Mid-Atlantic Ridge near 13° N containing a segment in which a number of linked detachment faults extend for 75 km along one flank of the spreading axis. The detachment faults are apparently all currently active and at various stages of development. A field of extinct core complexes extends away from the axis for at least 100 km. Our observations reveal the topographic characteristics of actively forming core complexes and their evolution from initiation within the axial valley floor to maturity and eventual inactivity. Within the surrounding region there is a strong correlation between detachment fault morphology at the ridge axis and high rates of hydroacoustically recorded earthquake seismicity. Preliminary examination of seismicity and seafloor morphology farther north along the Mid-Atlantic Ridge suggests that active detachment faulting is occurring in many segments and that detachment faulting is more important in the generation of ocean crust at this slow-spreading ridge than previously suspected.

The existence of low-angle faults extending deep below the axes of mid-ocean ridges^{7,8} has long been inferred from seafloor exposures of deep-seated rocks. Direct evidence for such faults has only recently come from bathymetry^{1,2} and seafloor sampling and drilling^{4,5,9}. The faults are corrugated parallel to the spreading direction and cap smooth topographic highs, termed oceanic core complexes, where deep-seated rocks such as gabbros and serpentinized peridotites are exposed^{4,5,10,11}. Most core complexes have been identified towards the ends of spreading segments where magma supply appears to be low, but in places they extend for tens of kilometres parallel to the axis^{4,5,12–14}. In parts of some segments, extension by low-angle faulting may have accounted for >50% of the total extension by spreading^{3,11,15}.

Because almost all core complexes identified to date are far enough from the axis to be inactive, the nature of active detachment faults is controversial. How do they initiate? How do they evolve as they emerge from the ocean floor? How is active detachment faulting accommodated along the length of the spreading axis? Is there a seismic signature to detachment faulting? We answer these questions using a new survey of the Mid-Atlantic Ridge (MAR) (D.K.S. & J.E., unpublished cruise report, RV *Knorr* Leg 182, June 2005) near 13° N together with existing multibeam bathymetry and hydroacoustically recorded seismicity, and extend our conclusions to other mid-ocean-ridge segments in the region.

A bathymetric map of the MAR between the Fifteen-Twenty and

the Marathon fracture zones (Fig. 1) shows an alternation in the morphology of the spreading axis. Of the two major segments in this area, that between 14° 35' N and 13° 50' N (labelled the 14° N segment) has a volcanic signature, with closely spaced volcanic ridges on both flanks, tens of kilometres long, and cut by steep inward-facing faults. The segment to the south between 13° 50' N and 12° 40' N (the 13° N segment) has more chaotic topography. The west flank, in particular, is distinctive, with widely spaced, narrow

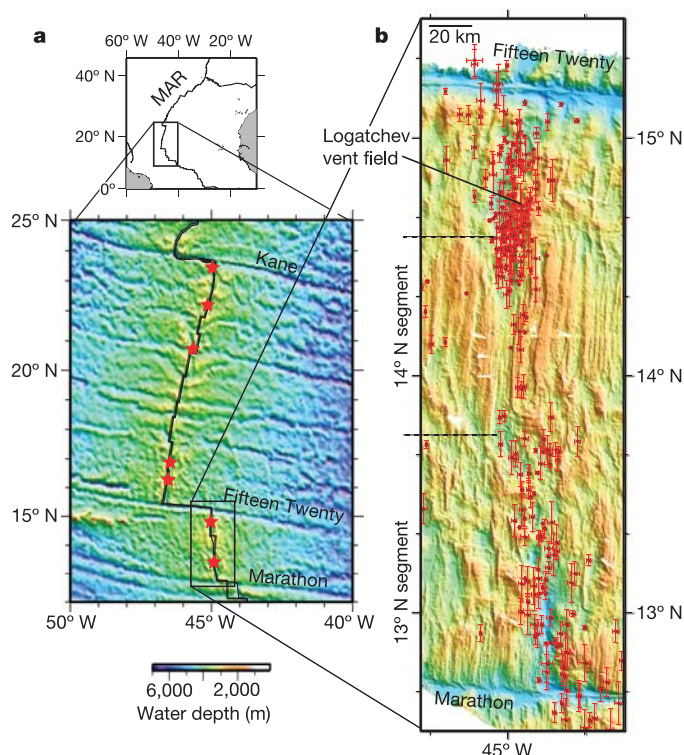


Figure 1 | Location map and bathymetry near 13° N on the MAR. **a**, Satellite altimetry data²⁵. The red stars indicate seven sections of the axis showing persistent, high levels of hydroacoustically recorded seismic activity²². The inset at the top shows the MAR axis. **b**, Multibeam bathymetry from refs 24 and 12 and D.K.S. & J.E., unpublished cruise report (RV *Knorr* Leg 182, June 2005). The '14° N segment' has faulted volcanic morphology, while the segments to the north and south have irregular and blocky topography and core complexes. The peridotite-hosted Logatchev hydrothermal vent field²⁶ is marked. Red dots show locations of 292 hydroacoustically detected events for the period 1999–2005 with 1σ error bars²² less than ±10 km.

¹Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA. ²School of Earth Sciences, University of Leeds, Leeds LS2 9JT, UK. ³CNRS, Marine Geosciences Group, IPGP Case 89, Paris 75252 cedex 05, France.

ridges, typically <20 km long, that have steep slopes dipping away from the spreading axis, and are separated by areas of smoother sea floor. As shown below, a large part of this smoother sea floor is corrugated parallel to the spreading direction.

In well-explored oceanic core complexes, the existence of a corrugated surface is correlated with detachment faulting and exposures of deep-seated rocks^{1–4,6}. Consequently, corrugations have been used to identify other core complexes in regions with limited data². Corrugations have a wavelength of hundreds of metres and an amplitude of tens of metres^{1,3}, and can be seen on shaded relief maps of multibeam bathymetry data. The origin of the corrugations is still enigmatic. Direct observations have shown that they are not produced by faults parallel to the spreading direction^{15,16}. One possibility is that they originate by continuous casting¹⁷ of a ductile footwall by irregularities in a strong and brittle hanging wall.

An inactive, corrugated core complex near 30° N (ref. 1) at the MAR (Fig. 2) shows other characteristics common to core complexes. First, a steep (25–30°) slope faces away from the axis at the older edge (scarps at slow- and intermediate-spreading ridges typically face towards the axis¹⁸). Second, a narrow, linear ridge parallel to the axis caps the outward-facing slope^{1,2} and extends beyond the corrugated surface. Dredge samples from the linear ridge in Fig. 2 show that it is composed of basalt¹ and thus is probably volcanic sea floor created at the axis. Finally, a steep normal fault dips towards the spreading axis, truncating the corrugated surface on its younger side.

In the 13° N segment, corrugated surfaces and outward facing steep slopes dominate the western flank of the spreading centre, extending for 75 km along axis and >100 km across axis (Fig. 3a). The feature centred at 13° 10' N, 45° 00' W (labelled '1' on Fig. 3a) is typical of many of those observed off-axis, and has a similar morphology to that shown in Fig. 2. The morphology and spatial

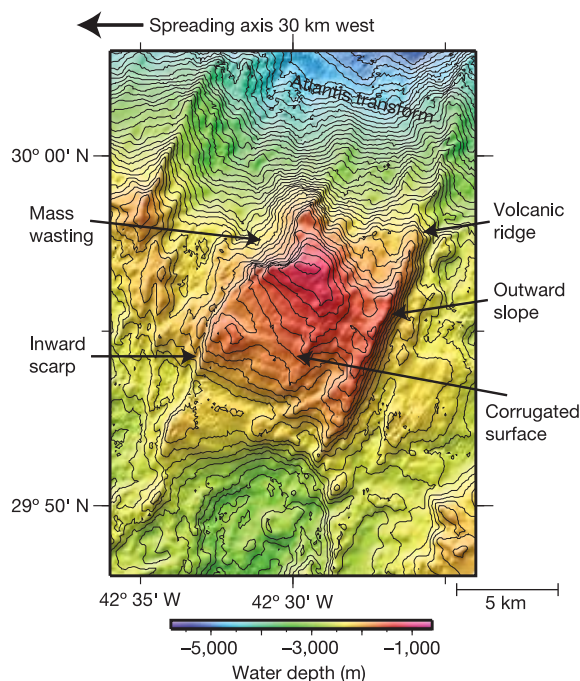


Figure 2 | Bathymetry map of a well-explored, extinct core complex south of the Atlantis fracture zone on the MAR³. Map contour interval is 100 m. The complex has spread ~30 km east of the axis. Its morphological characteristics include an outward-facing slope with a volcanically constructed ridge capping the core complex on the older (outer) side. A normal (inward-facing) fault scarp cuts the feature on its younger (inner) side. Corrugations running parallel to the spreading direction cap the shallow-dipping top. Serpentinized peridotite was recovered by dredging the corrugated surface⁷. Mass wasting of the massif is indicated by the scoop on its northwest corner.

scale of this and other corrugated surfaces and ridges in the 13° N segment indicate that the evolution of the western flank involves repeated detachment faulting and core complex formation.

Two core complexes north of complex '1' at 13° 20' N, 44° 55' W (labelled '2' on Fig. 3a, b) and 13° 30' N, 44° 55' W (labelled '3') have a different morphology from the core complexes that lie off-axis. Each extends ~10 km along-axis and is close to the spreading axis. The corrugated surfaces are domes that meet ridges to the west, and curve over to the east, dipping at 15° to meet the axial valley floor. Instead of being truncated by later faults, as with the off-axis core complexes, the slope at the east end of each of the domes intersects the median valley floor along a curved line <5 km from the volcanic axis (Fig. 3). These features suggest that the domes are the surfaces of detachment faults actively emerging from the valley floor.

We also identify a larger and apparently compound core complex extending for 30–40 km parallel to the axis (labelled 'C' on Fig. 3a, stretching from 12° 55' N to 13° 10' N). This compound core complex includes complex '1' and another similar complex to the south (labelled '4' in Fig. 3a) linked by a third complex located ~5 km to the east. This third complex terminates in a boundary with the volcanic morphology of the valley floor that is convex towards the east. A corrugated surface is present over a large part of the linking complex.

Because many of the core complexes in this segment have straight narrow ridges at their outer sides, we consider linear ridges as possible precursors to the emergence of the detachment faults. Such ridges occur in several places close to the spreading axis (labelled 'R' on Fig. 3a, b), each backed by a deep basin (Fig. 3). The outer slopes of the ridges dip at 15–20° away from the axis and show a hummocky volcanic morphology similar to those of the ridges farther from the spreading axis. Presumably they form by rotation of sections of volcanic sea floor away from the axis. We infer that the ridges mark the breakaway zones of detachment faults.

The strong morphologic evidence for active detachment faulting led us to search for associated seismic activity. Traditionally, seismically active MAR segments have been identified by locating foci of earthquakes of magnitude >4.5 detected by land seismometers, but their locations are poorly constrained, and these earthquakes are few in number. Between 1999 and 2005 an autonomous hydrophone array in the North Atlantic¹⁹ recorded the hydroacoustic energy from thousands of earthquakes of smaller magnitude (>2.5). Hydroacoustic events are better located than those detected teleseismically^{19–21}, and clearly show the currently active sections of MAR segments²². Two sections of ridge between the Fifteen-Twenty and Marathon fracture zones (Fig. 1b) show relatively high and persistent seismicity²³, and are separated by the seismically quiet 14° N segment. Core complexes have been identified in the active segment just south of the Fifteen-Twenty fracture zone²⁴. The other seismically active section, the 13° N segment, is active along its entire length, with no obvious spatial or temporal clustering of events. Location errors (Fig. 1b) are large enough, though, that events cannot be associated with individual detachments, nor with possible seismic sources associated with core complex formation.

The seismic evidence supports the morphologic evidence that a 75-km-long chain of detachment faults on the western side of the axial valley is active. Some of the detachment faults are mature (complexes '1' and '4' on Fig. 3a), some are in the early stages of development ('2' and '3'), and some are in an incipient stage of evolution, as shown by the narrow ridges close to the spreading axis ('R'). As they evolve and spread away from the axis, the detachments become inactive and new detachments initiate within a few kilometres of the axis.

There are a number of core complexes at different stages of evolution, so we can reconstruct the evolution of a core complex from initiation to maturity to inactivity (Fig. 4). The first stage is the subsidence of a basin within the axial valley floor, coupled with the emergence of a narrow basaltic ridge (Fig. 4a). The basin lies

immediately behind the ridge, and probably forms by outward footwall rotation as a new fault initiates. The inner (eastern) side of the basin (and hence the outer side of the ridge) has the volcanic morphology of the axial valley floor, tilted at 15° – 25° . As the ridge

evolves, the tilt of the outer slope increases to 20° – 30° . A fault scarp dipping at 15° – 25° emerges on the inner side of the ridge.

The next stage is the emergence of a domal corrugated fault surface that intersects the valley floor along a line convex towards the spreading axis (Fig. 4b). The corrugated surface dips at $\sim 15^{\circ}$ where it plunges into the valley floor. The dip of the surface gradually flattens as it emerges due to flexure of the footwall, and by ~ 5 km from its emergence the crest is nearly horizontal. Below the median valley floor, the detachment fault may continue at a low angle, but the early rotation of the footwall to form the basin suggests that the fault curves and steepens below the sea floor, so that the tip of the fault may lie several kilometres down.

The resulting mature core complex (Fig. 4c) becomes extinct when it is cut off by a normal fault. At that stage the domal surface flattens until it is close to horizontal. In some places, the core complex may extend as a single elongate unit from tens of kilometres to over a hundred kilometres from the spreading axis, as observed in the Australian–Antarctic Discordance^{10,11} and the Parece Vela basin¹³. In our study area, the small spacing between breakaway faults indicates a short life for any individual complex and regular nucleation of new detachment faults in the median valley.

Because of the apparent correlation of persistent earthquake seismicity and core complex morphology, we examined a larger region of the MAR (24° – 15° N) to look for additional sites of active core complex formation. We identified five ridge sections with high levels of hydroacoustically recorded seismic activity²² (Fig. 1a). The seismicity is persistent and not triggered by large earthquakes. Preliminary examination of available bathymetry from these and other segments in this larger area indicates a correlation between core complex morphology and seismicity. We suggest that detachment faulting may be more common than previously suspected in this part of the MAR. As much as 35% of the spreading axis may be experiencing detachment faulting and thus, $>15\%$ of the new seafloor accretion may be dominated by core complexes. We also suggest that the evolution of core complexes we have identified in our study area and the associated seismicity may be applicable to understanding other regions of active detachment faulting both in the

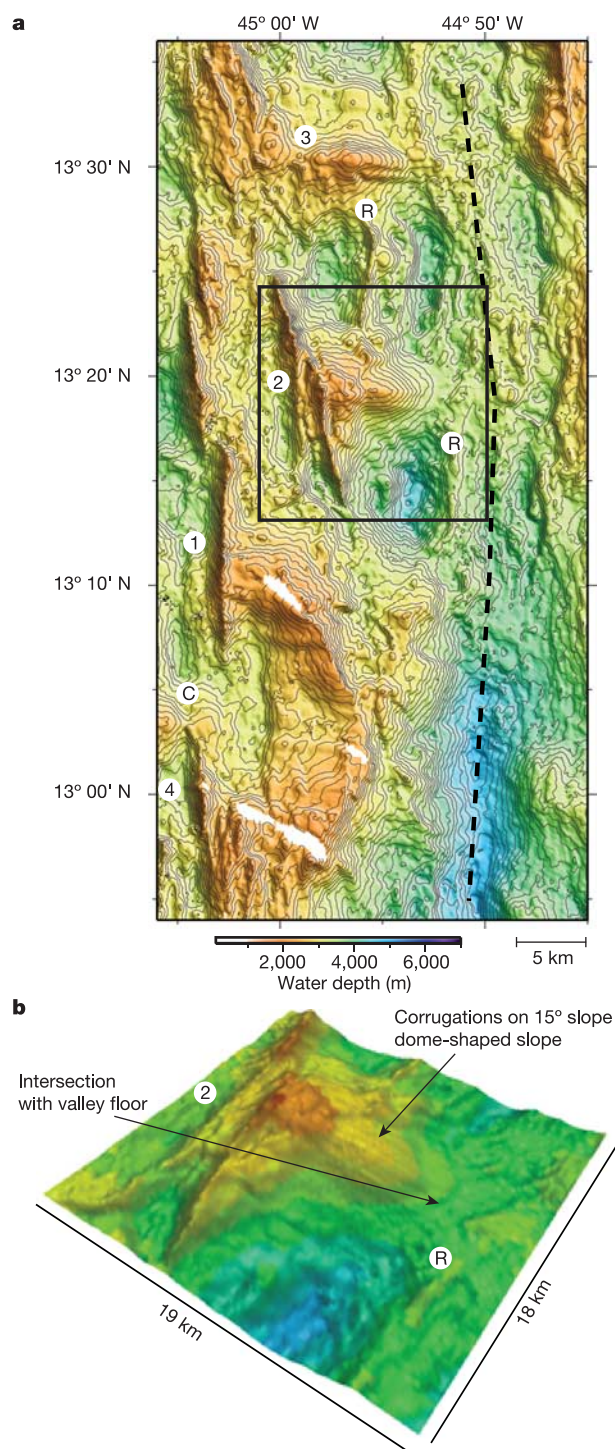


Figure 3 | Multibeam bathymetry showing detachment faults and core complexes in the 13° N segment. **a**, Map contour interval is 50 m. Linked and active core complexes extend ~ 75 km along the axis. The dashed line indicates the spreading axis. R, topographic ridges, inferred to be breakaways for new detachment surfaces. Numbers, complexes discussed in the text. C, compound core complex composed of complexes '1' and '4' and a third linking complex. **b**, Three-dimensional perspective view of complex '2' with no vertical exaggeration. The corrugated surface and its intersection with the sea floor are marked.

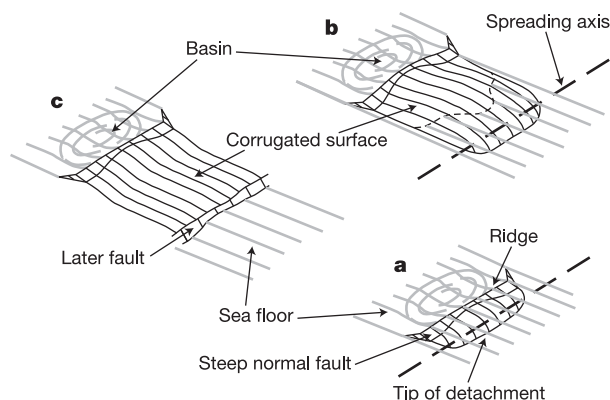


Figure 4 | Schematic three-dimensional oblique view of the evolution of a core complex by detachment faulting. Thick grey lines, sea floor (except for the detachment surface). Thin black lines, surface of the detachment fault and steeper normal faults, both exposed on the sea floor and below it. Thick dashed line, spreading axis. **a**, A breakaway ridge with a basin behind it. The basin is floored by volcanic sea floor, tilted up towards the breakaway ridge. The initially steep subsurface normal fault at the breakaway has already rotated to a shallower angle. A downward-curved detachment has started to form linked to the breakaway fault. Both faults are below the sea floor. **b**, The detachment has emerged. The thin dashed line marks the line of emergence. The basin is floored by volcanic sea floor, tilted up towards the breakaway ridge. The ridge has become arched, and little further rotation of the initial ridge has occurred. **c**, The detachment fault has been cut off by a later normal fault, perhaps of the breakaway of the next detachment. The dome has flattened, and the detachment fault is inactive.

oceans and on land, where the faults are more accessible but where erosion severely hinders their interpretation.

Received 9 April; accepted 2 June 2006.

1. Cann, J. R. *et al.* Corrugated slip surfaces formed at ridge–transform intersections on the Mid-Atlantic Ridge. *Nature* **385**, 329–332 (1997).
2. Tucholke, B. E., Lin, J. & Kleinrock, M. C. Megamullions and mullion structure defining oceanic metamorphic core complexes on the Mid-Atlantic Ridge. *J. Geophys. Res.* **103**, 9857–9866 (1998).
3. Blackman, D. K., Cann, J. R., Janssen, B. & Smith, D. K. Origin of extensional core complexes: Evidence from the Mid-Atlantic Ridge at Atlantis Fracture Zone. *J. Geophys. Res.* **103**, 21315–21333 (1998).
4. MacLeod, C. J. *et al.* Direct geological evidence for oceanic detachment faulting: The Mid-Atlantic Ridge, 15°45' N. *Geology* **30**, 279–282 (2002).
5. Escartín, J., Mével, C., MacLeod, C. J. & McCaig, A. M. Constraints on deformation conditions and the origin of oceanic detachments: The Mid-Atlantic Ridge core complex at 15°45' N. *Geochem. Geophys. Geosyst.* **4**, 1067, doi:10.1029/2002GC000472 (2003).
6. Searle, R. C. *et al.* FUJI Dome: A large detachment fault near 64 degrees E on the very slow-spreading Southwest Indian Ridge. *Geochem. Geophys. Geosyst.* **4**, doi:10.1029/2003GC000519 (2003).
7. Dick, H. J. B., Thompson, W. B. & Bryan, W. B. Low angle faulting and steady-state emplacement of plutonic rocks at ridge–transform intersections. *Eos* **62**, abstr. 406 (1981).
8. Karson, J. A. & Dick, H. J. B. Tectonics of ridge–transform intersections at the Kane Fracture Zone. *Mar. Geophys. Res.* **6**, 51–98 (1983).
9. Schroeder, T. & John, B. E. Strain localization on an oceanic detachment fault system, Atlantis Massif, 30 degrees N, Mid-Atlantic Ridge. *Geochem. Geophys. Geosyst.* **5**, 1–30 (2004).
10. Christie, D. M., West, B. P., Pyle, D. G. & Hanan, B. B. Chaotic topography, mantle flow and mantle migration in the Australian–Antarctic Discordance. *Nature* **394**, 637–644 (1998).
11. Okino, K., Matsuda, K., Christie, D., Nogi, Y. & Koizumi, K. Development of oceanic detachment and asymmetric spreading at the Australian–Antarctic Discordance. *Geochem. Geophys. Geosyst.* **5**, doi:10.1029/2004GC000793 (2004).
12. Escartín, J. & Cannat, M. Ultramafic exposures and the gravity signature of the lithosphere near the Fifteen–Twenty Fracture Zone (Mid-Atlantic Ridge, 14°–16.5°N). *Earth Planet. Sci. Lett.* **171**, 411–424 (1999).
13. Ohara, Y., Yoshida, T. & Kasuga, S. Giant megamullion in the Perece Vela Backarc basin. *Mar. Geophys. Res.* **22**, 47–61 (2001).
14. Cannat, M. *et al.* Thin crust, ultramafic exposures, and rugged faulting patterns at the Mid-Atlantic Ridge (22°–24°N). *Geology* **23**, 49–52 (1995).
15. Blackman, D. K. *et al.* Geology of the Atlantis Massif (Mid-Atlantic Ridge, 30°N): Implications for the Evolution of an Ultramafic Oceanic Core Complex. *Mar. Geophys. Res.* **23**, 443–460 (2003).
16. Tucholke, B. E., Fujioka, K., Ishihara, T., Hirth, G. & Kinoshita, M. Submersible study of an oceanic megamullion in the central North Atlantic. *J. Geophys. Res.* **106**, 16145–16161 (2001).
17. Spencer, J. E. Geologic continuous casting below continental and deep-sea detachment faults and at the striated extrusion of Sacsayhuaman, Peru. *Geology* **27**, 327–330 (1999).
18. Macdonald, K. C. Mid-ocean ridges: Fine scale tectonic, volcanic, and hydrothermal processes within the plate boundary zone. *Annu. Rev. Earth Planet. Sci.* **10**, 155–190 (1982).
19. Smith, D. K. *et al.* Hydroacoustic monitoring of seismicity at the slow-spreading Mid-Atlantic Ridge. *Geophys. Res. Lett.* **29**, doi:10.1029/2001GL013912 (2002).
20. Bohnenstiehl, D. R., Tolstoy, M., Dziak, R. P., Fox, C. G. & Smith, D. K. Aftershock sequences in the mid-ocean ridge environment: An analysis using hydroacoustic data. *Tectonophysics* **354**, 49–70 (2002).
21. Fox, C. G., Matsumoto, H. & Lau, T.-K. Monitoring Pacific ocean seismicity from an autonomous hydrophone array. *J. Geophys. Res.* **106**, 4183–4206 (2001).
22. Smith, D. K. *et al.* Spatial and temporal distribution of seismicity along the northern Mid-Atlantic Ridge (15°–35°N). *J. Geophys. Res.* **108**, doi:10.1029/2002JB001964 (2003).
23. Escartín, J., Smith, D. K. & Cannat, M. Parallel bands of seismicity at the Mid-Atlantic Ridge, 12–14°N. *Geophys. Res. Lett.* **30**, doi:10.1029/2003GL017226 (2003).
24. Fujiwara, T. *et al.* Crustal evolution of the Mid-Atlantic Ridge near the Fifteen–Twenty Fracture Zone in the last 5 Ma. *Geochem. Geophys. Geosyst.* **4**, 1024, doi:10.1029/2002GC000364 (2003).
25. Smith, W. H. F. & Sandwell, D. T. Global sea floor topography from satellite altimetry and ship depth soundings. *Science* **277**, 1956–1962 (1997).
26. Sudarikov, S. M. & Roumiantsev, A. B. Structure of hydrothermal plumes at the Logatchev vent field, 14 45' N, Mid-Atlantic Ridge: evidence from geochemical and geophysical data. *J. Vol. Geotherm. Res.* **101**, 245–252 (2000).

Acknowledgements We are grateful to the captain and crew of the RV *Knorr* (Leg 182, 2005). We are also grateful to P. Lemmond, H. Schouten and M. Tivey for their help in collecting the data at 13° N. We had fruitful discussions with R. Searle, H. Schouten, M. Tivey and R. Sohn. We also thank J. Goff for constructive comments on the manuscript. This work was supported by the National Science Foundation. This is an IGP contribution.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.K.S. (dsmith@whoi.edu).

LETTERS

A ubiquitous thermoacidophilic archaeon from deep-sea hydrothermal vents

Anna-Louise Reysenbach¹, Yitai Liu¹, Amy B. Banta¹, Terry J. Beveridge², Julie D. Kirshtein³, Stefan Schouten⁴, Margaret K. Tivey⁵, Karen L. Von Damm⁶ & Mary A. Voytek³

Deep-sea hydrothermal vents are important in global biogeochemical cycles, providing biological oases at the sea floor that are supported by the thermal and chemical flux from the Earth's interior. As hot, acidic and reduced hydrothermal fluids mix with cold, alkaline and oxygenated sea water, minerals precipitate to form porous sulphide–sulphate deposits. These structures provide microhabitats for a diversity of prokaryotes that exploit the geochemical and physical gradients in this dynamic ecosystem¹. It has been proposed that fluid pH in the actively venting sulphide structures is generally low (pH < 4.5)², yet no extreme thermoacidophile has been isolated from vent deposits. Culture-independent surveys based on ribosomal RNA genes from deep-sea hydrothermal deposits have identified a widespread euryarchaeotal lineage, DHVE2 (deep-sea hydrothermal vent euryarchaeotic 2)^{3–6}. Despite the ubiquity and apparent deep-sea endemism of DHVE2, cultivation of this group has been unsuccessful and thus its metabolism remains a mystery. Here we report the isolation and cultivation of a member of the DHVE2 group, which is an obligate thermoacidophilic sulphur- or iron-reducing heterotroph capable of growing from pH 3.3 to 5.8 and between 55 and 75 °C. In addition, we demonstrate that this isolate constitutes up to 15% of the archaeal population, providing evidence that thermoacidophiles may be key players in the sulphur and iron cycling at deep-sea vents.

Since their discovery in 1977, deep-sea hydrothermal vents have provided one of the few environments on Earth for searching for the upper temperature limits of life; these ecosystems may represent models for both the origin of life on Earth and exploration of life on other planets. This ecosystem, fuelled largely by geochemical energy, hosts many invertebrates new to science, and they often thrive because of their endosymbiotic bacterial partners. Likewise, these deep-sea oases have provided a vast array of newly described free-living microbes, many associated with the actively forming porous deep-sea vent deposits such as 'chimneys'. Here, the contrasting geochemistry of hydrothermal fluids and sea water, the mineralogy of porous substrates, and microbial activity are tightly coupled.

The steep chemical and thermal gradients within the walls of these vent deposits provide a wide range of microhabitats for microorganisms⁷. For example, calculations of transport across chimney walls² showed that the pH of fluids (at <120 °C) within pores of chimneys should be very low (pH < 3) if only diffusion is occurring. If some advection of sea water into the chimney occurs, the porewater pH at 50–80 °C should still be low (around pH 4–6); with advection of vent fluid outward, pH at 50–80 °C should range from <3 to 6 for low rates of flow, or from pH 3 to 4.5 (similar to that of the vent fluid measured at 25 °C) for rapid rates of flow^{2,8}. This agrees well with an

in situ measurement of pH made at the exterior of a chimney wall at 13°N on the East Pacific Rise (EPR), beneath a colony of *Alvinella pompejana*; the pH was 4.1 at a temperature of 120 °C (ref. 9). In addition, localized acidity is produced chemically as minerals such as pyrite and chalcopyrite form: $\text{Fe}^{2+} + 2\text{H}_2\text{S} = \text{FeS}_2 + \text{H}_2 + 2\text{H}^+$ and $\text{Cu}^+ + \text{Fe}^{2+} + 2\text{H}_2\text{S} = \text{CuFeS}_2 + 0.5\text{H}_2(\text{aq.}) + 3\text{H}^+$. It is therefore surprising that all thermophiles thus far isolated from deep-sea vents are neutrophiles, or slightly acidophilic¹⁰ and at best acidotolerant¹¹, and grow very poorly or not at all at the predicted low pH conditions from which they were isolated. For example, *Thermococcus*, perhaps the most frequently cultivated thermophile from deep-sea vents, grows best around pH 6.5, with the majority of species unable to grow below pH 5.5 and only some species can grow very poorly at about pH 4 (ref. 12).

Culture-dependent and culture-independent approaches have exposed a vast diversity of Bacteria and Archaea associated with deep-sea vent deposits (see, for example, refs 1, 13, 14), with the Bacteria being dominated by many novel branches of the ϵ -Proteobacteria¹⁵. However, although numerous Archaea have been isolated from these deposits, few isolates are found in environmental 16S rRNA gene clone libraries and most environmental clones have no representatives in culture. In particular, one lineage is widespread at deep-sea vents and is frequently associated with actively venting sulphide deposits. The 'deep-sea hydrothermal vent euryarchaeotic' lineage DHVE2 is often found in clone libraries (see, for example, refs 3–6, 14, 16) and has been reported as the most dominant clone type in some archaeal clone libraries⁶ (Supplementary Tables 1 and 2).

Using a specific primer set for the DHVE2 16S rRNA gene (Supplementary Methods), we screened 63 samples from multiple vent sites along two ridge systems (the EPR from 9° 16.8' N, 104° 13' W to 20° 50' N, 109° 07' W, and the Eastern Lau Spreading Centre, ELSC, from 20° 3.2' S, 176° 8.0' W to 22° 10.8' S, 176° 36.1' W) during three different research cruises (Supplementary Table 2, Supplementary Fig. S1). The DHVE2 were identified in 36 samples, and represented up to 10% and 15% of the archaeal population at EPR and ELSC, respectively (as determined by quantitative polymerase chain reaction (qPCR); Table 1). Generally, the occurrence of DHVE2 appeared to be associated with the fine-grained white elemental sulphur (S₈; identified in three different samples by X-ray diffraction; Supplementary Methods) or iron oxyhydroxide (X-ray amorphous—identified on the basis of orange colour and texture; Supplementary Table 3) deposits that often coat the outside of mature actively venting sulphide deposits (Supplementary Fig. S2). Additionally, owing to the high guanine and cytosine content of the 16S rRNA sequence and its occasional co-occurrence with thermophiles such as *Thermococcus*, it has been

¹Department of Biology, Portland State University, Portland, Oregon 97201, USA. ²Department of Molecular & Cellular Biology, University of Guelph, Guelph, Canada.

³US Geological Survey, MS 430, 12201 Sunrise Valley Drive, Reston, Virginia 20192, USA. ⁴Department of Marine Biogeochemistry & Toxicology, Royal Netherlands Institute for Sea Research, 1790 AB Den Burg, Texel, The Netherlands. ⁵Department of Marine Chemistry and Geochemistry, WHOI, Woods Hole, Massachusetts 02543, USA. ⁶Complex Systems Research Center, EOS, University of New Hampshire, 366 Morse Hall, 39 College Road, Durham, New Hampshire 03824-3525, USA.

Table 1 | Relative DHVE2 abundance at various sites

Site	Range (%)	Average (%)	Median (%)
Kilo Moana (n = 3)*	0.08–6.21	2.16	0.2
Tow Cam (n = 5)*	0.5–14.8	4.35	2.11
ABE (n = 3)*	0.5–5.3	3.19	3.78
Tui Malila (n = 8)*	0.1–13.2	3.96	2.87
Mariner (n = 6)*	0.01–5.0	1.28	0.26
ELSC total (n = 25)*	0.01–15	2.98	1.20
EPR (n = 4)*	0.75–10	3.28	1.37

Percentage abundance of DHVE2 as compared to total Archaea (DHVE2/Archaea) at sites along two ridge systems, the East Pacific Rise (EPR, 9° 50.3' N, 104° 17.5' W to 9° 33.5' N, 104° 14.9' W) and the Eastern Lau Spreading Centre (ELSC, 20° 3.2' S, 176° 8' W to 22° 10.8' S, 176° 36.1' W; Kilo Moana, Tow Cam, ABE, Tui Malila, Mariner).
* Number of samples analysed. qPCR of each sample was run in duplicate.

suggested that the DHVE2 are thermophiles, perhaps heterotrophs whose growth may be stimulated by sulphur^{1,16}. A DHVE2 genome fragment from a metagenome library contained a thermostable DNA polymerase, which was further suggestive of the thermophilic nature of this lineage¹⁷. However, all attempts to culture this group—using standard media and conditions to select for sulphate reducers, methanogens or fermenters at 60, 80 and 90 °C and around pH 6.5—have been unsuccessful⁵.

On the basis of predictions that conditions within many sulphide deposits should be acidic, and owing to the presence of novel sequences most closely related to the acidophiles *Thermoplasma*, *Picrophilus* and *Ferroplasma* (Fig. 1, sequence LART667E06) and the presence of the DHVE2 group from two samples obtained from the Mariner deep-sea vents along the ELSC (22° 10.8' S, 176° 36.1' W—depth of vent field ~1,920 m; ref. 18), we enriched for heterotrophic thermoacidophiles (pH 4.5) using trypticase peptone, yeast extract and elemental sulphur. After a week, two cultures (designated strains T449 and T469) were purified by three serial dilutions. Comparative sequence analysis of the 16S rRNA gene of strains T449 and T469 placed these isolates within the DHVE2. The group forms a monophyletic clade, and isolates are closely related (about 95–97% similar) to cloned 16S rRNA gene sequences obtained from deep-sea vents and a sequence obtained from methane-hydrate-bearing marine

Table 2 | Phenotypic characteristics of the T469 and its closest thermophilic relatives

	T469	<i>Thermoplasma</i> spp. ²⁹	<i>Picrophilus</i> spp. ²³
Morphology	Pleiomorphic cocci	Pleiomorphic cocci	Pleiomorphic cocci
Flagella	Yes	No	No
S-layer present	Yes	No	Yes
Metabolism	O	O, M	O
Electron donor	Yeast extract, casein, trypticase peptone	Yeast extract, meat extract, some sugars	Yeast extract, tryptone
Electron acceptor	Fe ³⁺ , S ⁰	Complex organic carbon, S ⁰	Complex organic carbon, O ₂
Aerobic growth	No	Yes	Yes
Anaerobic growth	Yes	Yes	No
Marine/terrestrial	Marine	Marine and terrestrial	Terrestrial
Temp. range (°C)	55–75; 70*	33–67; 60*	47–65; 60*
pH range	3.3–5.8; 4.5*	0.5–4; 2*	0–3.5; 0.7*

O, chemoorganotroph (heterotroph); M, chemomixotroph.
* Optimum.

sediments at about 200 m below the sea floor (m.b.s.f.) from the Cascadia margin¹⁹ (Supplementary Table 1). The new isolates are only 83% similar in 16S rRNA sequence to their closest relative in culture, *Thermoplasma volcanium*, an isolate from shallow marine vents and solfataras²⁰ that grows best at pH 2.0 and 60 °C (Table 2). Additional thermoacidophilic enrichments of DHVE2 were obtained from two sites along the ELSC (Kilo Moana, 20° 3.2' S, 176° 8.0' W, ~2,620 m, and Tow Cam, 20° 19.1' S, 176° 8.2' W, ~2,720 m; ref. 21) and five sites along the EPR (at 9° 17' N, 9° 33' N, 9° 46' N, 9° 47' N, 9° 50' N). One strain from the Mariner vent site, T469, was further characterized and its purity was confirmed by repeated sequencing of the 16S rRNA gene.

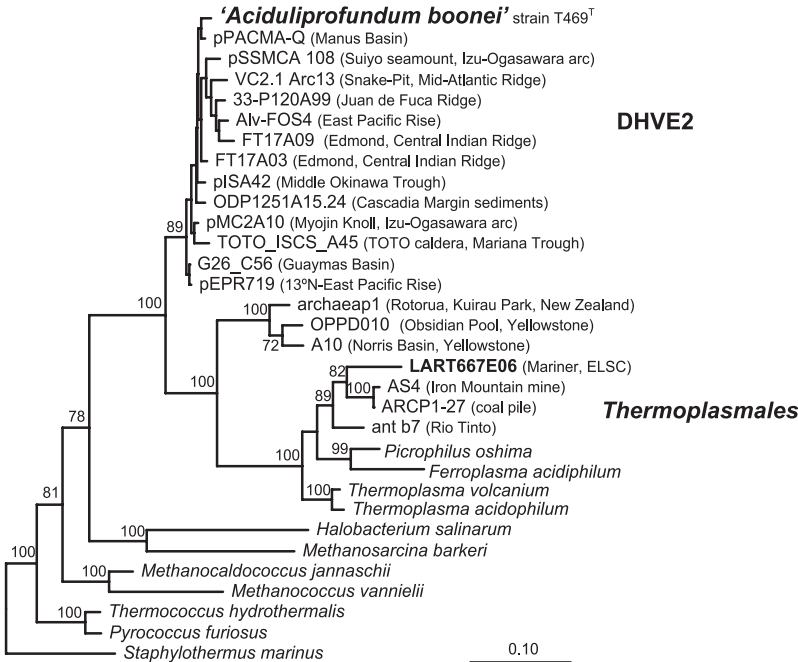


Figure 1 | Phylogenetic relationships between 16S rRNA sequences from strain T469 and representatives of the DHVE2 and Euryarchaeota. The tree was inferred using maximum-likelihood analysis. The scale bar

represents 0.1 substitutions per nucleotide position. The numbers at the branch nodes are bootstrap values (>70%) based on 100 bootstrap resamplings. The sequence LART667E066 (bold) is discussed in the text.

Strain T469 is, to our knowledge, the first obligate thermoacidophile to be reported from deep-sea vents and the first cultivated member of the DHVE2 group. The organism grows between pH 3.3 and 5.8, growing best between pH 4.2 and 4.8 (Supplementary Fig. S3) at 70 °C and to a maximum cell density of about $(4\text{--}6) \times 10^8$ cells ml⁻¹. The isolate does not grow above 77 °C or below 50 °C, but like *Thermococcales* and *Thermotogales* it is an obligate anaerobe requiring elemental sulphur (or ferric iron), NaCl (optimum 2.5–3.5% w/v) and complex organics such as trypticase peptone, casein and yeast extract for growth (Supplementary Methods, Table 2). Chemical analyses showed that its membrane lipids are predominantly composed of glycerol dibiphytanyl glycerol tetraethers containing 0–4 cyclopentane rings. These membrane lipids have also been detected in the closest cultured relatives *Ferropasma* and *Thermoplasma* and are thought to be unique to acidophilic Crenarchaeota and Euryarchaeota²².

The cells are pleiomorphic to spherical, about 0.6–1 µm in diameter and are motile with a single flagellum (Fig. 2a, b). Some flagella have their proximal end encased in an unusual complex periodic sheath (Fig. 2c). Like many other Archaea (for example, *Methanococcus* and *Sulfolobus*), they are enveloped by a plasma membrane and a single S-layer (Fig. 2a, e). The S-layer, similar to that of *Picrophilus oshimae*²³, is thick (~40 nm), with a presumed tetragonal (p4) lattice and resembles a delicate bridal veil bordering the cells (Fig. 2a). Although S-layers are quasi-crystalline, this S-layer bends into small highly curved structures (Fig. 2d, vesicles). These vesicles bud from the cell, partitioning small quantities of cytoplasm that could then anneal to adjacent cells. This is a process that is common among their Gram-negative bacterial counterparts²⁴. The ability of DHVE2's S-layer to bend into such high curvature structures as these vesicles suggests that the bonding forces between S-layer subunits are weak or transient. This is unusual²⁵, especially as this S-layer is the sole cell wall structure.

As this organism is the first isolate in the DHVE2 clade, representing a new Order within the phylum Euryarchaeota, we propose the following candidate status of this taxon;

'Aciduliprofundum boonei', gen. et sp. nov.

Etymology. Acidulus (Latin), a little sour, acidulous; profundus (Latin), deep; *Aciduliprofundum*, a bacterium isolated from acidic deep-sea vents; boonei (Latin), of Boone, named in honour of David Boone for his contributions to the study of archaeal diversity.

Diagnosis. Anaerobic heterotrophic sulphur- and iron-reducing thermoacidophile belonging to the kingdom Euryarchaeota. Cells are pleiomorphic to spherical, about 0.6–1 µm in diameter, with a single flagellum and an S-layer. Grows best at 70 °C and pH 4.2–4.8. From actively venting deep-sea vent deposits, first isolated from deep-sea vents at 22° 10.8' S, 176° 36.1' W (Mariner vent field).

The successful cultivation of this thermoacidophile from deep-sea vents points to the importance of the role that culture-independent approaches have played in exposing the vastness and ecological significance of the diversity of microbes, but also to the integrative approach where geochemical modelling has predicted the environmental conditions in which this diversity thrives. Furthermore, this archaeon shares an overlapping niche with other fast-growing thermophilic heterotrophs, such as *Thermococcus*. It is likely that under most enrichment conditions of higher pH, *Thermococcus* would rapidly outcompete the DHVE2, and this prevented our successful isolation of this novel lineage in the past. The DHVE2 are often found co-occurring with lineages such as the ε-Proteobacteria (based on 16S rRNA gene data, not shown). These bacteria are probably autotrophic sulphide/sulphur or hydrogen oxidizers¹⁵ that could create a localized anaerobic environment (and also biologically lower the local pH), providing a microniche for anaerobic acidophiles. Additionally, owing to the association of DHVE2 with dense biofilms that form among iron oxides and elemental sulphur, it is possible that these communities provide both the oxidized substrates and the complex organics to support the observed metabolism of this archaeon.

With this confirmation of extreme thermoacidophiles from deep-sea vents, it is likely that this physiological group plays a central role in the biogeochemistry and mineralogy at deep-sea vents. It is thus possible that the numerous lineages that have been identified from similar environments using culture-independent approaches may be acidophiles, or inhabit some microhabitat yet to be identified. Such isolates will not only expand our understanding of the role these organisms play in the hydrothermal vent environment, but also provide clues to the biosignatures that remain in the ancient hydrothermal rock record.

METHODS

Full details are given in the Supplementary Methods.

Sampling sites. Actively venting sulphide samples were collected using the DSRV *Alvin* or DSOV *Jason II* from the EPR or the ELSC, respectively. Once on board ship, subsamples where the DHVE2 were detected were taken by using sterile spatulas and removing the ~1–3-mm-deep superficial sulphur or iron crusts. Samples were stored anaerobically in serum vials at 4 °C as rock slurries or frozen immediately for DNA extractions.

Enrichment cultivation and isolation. Enrichment cultures were started several months later in the land-based laboratory at Portland State University. Initial enrichments were incubated at 70 °C in anaerobic marine media (Supplementary Methods). Growth was monitored by cell counts in triplicate cultures, and repeated at least twice. A suite of carbon source sugars and proteinaceous compounds were tested at 0.1–0.5% (w/v, final concentration; Supplementary Methods) as alternative carbon sources in media amended with vitamins or 0.01% yeast extract.

Quantitative PCR. Samples were analysed for DHVE2 abundance and total archaeal abundance by quantitative PCR (qPCR; Supplementary Methods).

Phylogenetic analysis of 16S rRNA sequences. 16S rRNA gene sequences were amplified by PCR (Supplementary Methods) and manually aligned on the basis of secondary structure constraints in the ARB software package²⁶. Phylogenetic relationships were determined using evolutionary distance and maximum likelihood analysis. Phylogenetic trees were constructed using 1,356 homologous nucleotides with the fastDNAmI program²⁷, using an optimized T transition/transversion ratio of 1.52 and the global branch rearrangements option.

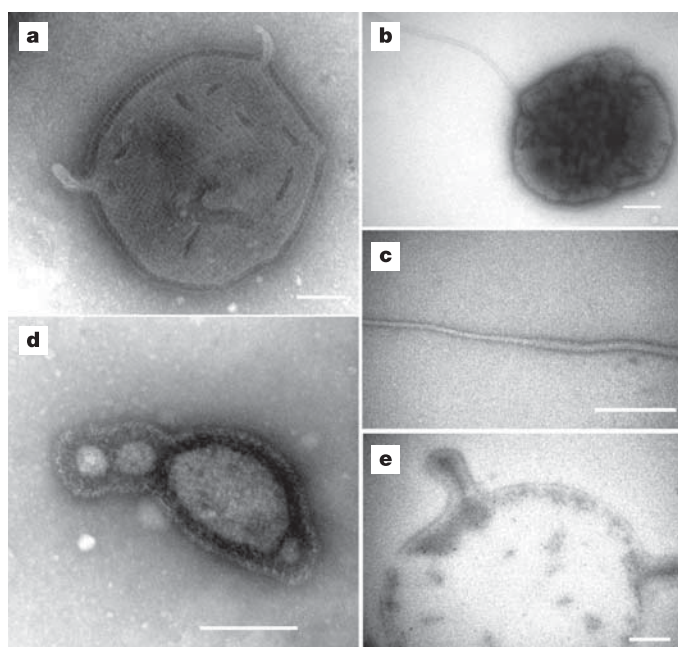


Figure 2 | Electron photomicrographs of T469. **a**, A negatively stained electron micrograph image showing the S-layer and the initiation of vesicle formation. **b**, Negatively stained electron micrograph image showing the flagellum. **c**, Detail of the flagellum with thicker sheath area on right. **d**, Multiple vesicles with the thick S-layer. **e**, A thin section of a cell showing the bending of the S-layer. Scale bar in all panels, 200 nm.

Confidence of tree topologies were estimated by performing 100 bootstrap replicates using fastDNAmL.

Membrane lipid analysis. Membrane lipids were extracted (Supplementary Methods) and analysed by high performance liquid chromatography/atmospheric pressure chemical ionization/mass spectrometry²⁸.

Transmission electron microscopy. Both negatively stained cells and stained thin sections (Supplementary Methods) were viewed with a Philips CM10 operating under standard conditions at 100 kV.

Received 24 April; accepted 19 May 2006.

- Reysenbach, A. L. & Shock, E. L. Merging genomes with geochemistry in hydrothermal ecosystems. *Science* **296**, 1077–1082 (2002).
- Tivey, M. K. in *Subseafloor Biosphere at Mid-Ocean Ridges* (eds Wilcock, W., Cary, C., DeLong, E., Kelley, D. & Baross, J.) 137–152 (Geophysical Monograph Series, American Geophysical Union, Washington DC, 2004).
- Takai, K. & Horikoshi, K. Genetic diversity of archaea in deep-sea hydrothermal vent environments. *Genetics* **152**, 1285–1297 (1999).
- Reysenbach, A.-L., Longnecker, K. & Kirshtein, J. Novel bacterial and archaeal lineages from an *in situ* growth chamber deployed at a Mid-Atlantic Ridge hydrothermal vent. *Appl. Environ. Microbiol.* **66**, 3798–3806 (2000).
- Nercessian, O., Reysenbach, A.-L., Prieur, D. & Jeanthon, C. Compositional changes in archaeal communities associated with deep-sea hydrothermal vent samples from the East Pacific Rise (13°N). *Environ. Microbiol.* **5**, 492–502 (2003).
- Hoek, J., Banta, A. B., Hubler, R. F. & Reysenbach, A.-L. Microbial diversity of a deep-sea hydrothermal vent biofilm from Edmond vent field on the Central Indian Ridge. *Geobiology* **1**, 119–127 (2003).
- McCollom, T. M. & Shock, E. L. Geochemical constraints on chemolithoautotrophic metabolism by microorganisms in seafloor hydrothermal systems. *Geochim. Cosmochim. Acta* **61**, 4375–4391 (1997).
- Von Damm, K. L. in *Physical, Chemical, Biological, and Geological Interactions within Seafloor Hydrothermal Systems* (eds Humphris, S., Lupton, J., Mullineaux, L. & Zierenberg, R.) 222–247 (American Geophysical Union, Washington DC, 1995).
- Le Bris, N., Zbinden, M. & Gaill, F. Processes controlling the physicochemical micro-environments associated with Pompeii worms. *Deep Sea Res.* **52**, 1071–1083 (2005).
- Takai, K. *et al.* *Lebetimonas acidiphila* gen. nov., sp. nov., a novel thermophilic, acidophilic, hydrogen-oxidizing chemolithoautotroph within the 'Epsilonproteobacteria', isolated from a deep-sea hydrothermal fumarole in the Mariana Arc. *Int. J. Syst. Evol. Microbiol.* **55**, 183–189 (2005).
- Prokofeva, M. *et al.* Cultivated anaerobic acidophilic/acidotolerant thermophiles from terrestrial and deep-sea hydrothermal habitats. *Extremophiles* **9**, 437–448 (2005).
- Kobayashi, T. in *Bergeys Manual of Systematic Bacteriology* (eds Boone, D. R., Castenholz, R. W. & Garrity, G. M.) 342–346 (Springer, New York, 2001).
- Schrenk, M. O., Kelley, D. S., Delaney, J. R. & Baross, J. A. Incidence and diversity of microorganisms within the walls of an active deep-sea sulfide chimney. *Appl. Environ. Microbiol.* **69**, 3580–3592 (2003).
- Nakagawa, T. *et al.* Geomicrobiological exploration and characterization of a novel deep-sea hydrothermal system at the TOTO caldera in the Mariana volcanic arc. *Environ. Microbiol.* **8**, 37–49 (2006).
- Nakagawa, S. *et al.* Distribution, phylogenetic diversity and physiological characteristics of epsilon-Proteobacteria in a deep-sea hydrothermal field. *Environ. Microbiol.* **7**, 1619–1632 (2005).
- Takai, K., Komatsu, T., Inagari, F. & Horikoshi, K. Distribution of Archaea in a black smoker chimney structure. *Appl. Environ. Microbiol.* **67**, 3618–3629 (2001).
- Moussard, H. *et al.* Thermophilic lifestyle for an uncultured archaeon from hydrothermal vents: evidence from environmental genomics. *Appl. Environ. Microbiol.* **72**, 2268–2271 (2006).
- Ishibashi, J. *et al.* Expedition reveals changes in Lau Basin hydrothermal system. *Eos* **87**, 13–17 (2006).
- Inagaki, F. *et al.* Biogeographical distribution and diversity of microbes in methane hydrate-bearing deep marine sediments on the Pacific Ocean Margin. *Proc. Natl Acad. Sci. USA* **103**, 2815–2820 (2006).
- Seeger, A., Langworthy, T. A. & Stetter, K. O. *Thermoplasma acidophilum* and *Thermoplasma volcanium* sp. nov. from solfatara fields. *Syst. Appl. Microbiol.* **10**, 161–171 (1988).
- Langmuir, C. H. *et al.* Hydrothermal prospecting and petrological sampling in the Lau Basin: Background data for the integrated study site. *Eos* **85**, B13A–0189 (2004).
- Macalady, J. L. *et al.* Membrane monolayers in *Ferroplasma* spp.: a key to survival in acid. *Extremophiles* **8**, 411–419 (2004).
- Schleper, C. *et al.* *Picrophilus* gen. nov., fam. nov.: a novel aerobic, heterotrophic, thermoacidophilic genus and family comprising Archaea capable of growth around pH 0. *J. Bacteriol.* **177**, 7050–7059 (1995).
- Beveridge, T. J. Structures of Gram-negative cell walls and their derived membrane vesicles. *J. Bacteriol.* **7**, 253–260 (1999).
- Sleytr, U. B. & Beveridge, T. J. Bacterial S-layers. *Trends Microbiol.* **7**, 253–260 (1999).
- Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004).
- Olsen, G., Matsuda, H., Hagstrom, R. & Overbeek, R. FastDNAmL: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Comput. Appl. Biosci.* **10**, 41–48 (1994).
- Hopmans, E. C. *et al.* Analysis of intact tetraether lipids in archaeal cell material and sediments by high performance liquid chromatography/atmospheric pressure chemical ionization mass spectrometry. *Rapid Comm. Mass. Spectrom.* **14**, 585–589 (2000).
- Seeger, A., Langworthy, T. A. & Stetter, K. O. *Thermoplasma acidophilum* and *Thermoplasma volcanium* sp. nov. from solfatara fields. *Syst. Appl. Microbiol.* **10**, 161–171 (1988).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the crew of the RV *Melville*, RV *Atlantis* and the DSROV *Jason II* and DSV *Alvin* teams for their assistance. This work is funded by the US National Science Foundation (A.-L.R., M.K.T., K.L.V.D.), a PSU Faculty Enhancement Award (A.-L.R.), the Natural Science and Engineering Research Council of Canada (NSERC), the US Department of Energy (T.J.B.), the US National Research Program, Water Resources Division, USGS, and NASA Exobiology (M.A.V.). We thank D. Moyle and R. Harris for technical help with TEM; P. Craddock for XRD; M. Baas and E. Hopmans for analytical assistance with the lipid analyses; I. Cozzarelli, M. Doughten and J. Jaeschke for the organic acid analysis; G. Sincerny, A. O'Neill and D. Decker for their assistance; and I. Ferrera for sharing qPCR results. J. Euzéby and M. Bartlett are acknowledged for assistance with naming the new archaeon.

Author Contributions A.-L.R. and Y.L. isolated and characterized the acidophile; A.B.B. did all molecular phylogenetic and environmental DNA analyses; M.A.V. and J.D.K. did the qPCR analysis; T.J.B. did the ultrastructure analysis and interpretation; S.S. performed membrane lipid analysis; M.K.T. provided XRD data and geochemical interpretation; and K.L.V.D. provided samples, data on the sample locations and geochemical context. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The new sequences described in this manuscript have been deposited in GenBank, accession numbers DQ451875 (T469) and DQ451876 (T449). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.-L.R. (reysenbacha@pdx.edu).

LETTERS

Strategies for mitigating an influenza pandemic

Neil M. Ferguson¹, Derek A. T. Cummings², Christophe Fraser¹, James C. Cajka³, Philip C. Cooley³
& Donald S. Burke²

Development of strategies for mitigating the severity of a new influenza pandemic is now a top global public health priority. Influenza prevention and containment strategies can be considered under the broad categories of antiviral, vaccine and non-pharmaceutical (case isolation, household quarantine, school or workplace closure, restrictions on travel) measures¹. Mathematical models are powerful tools for exploring this complex landscape of intervention strategies and quantifying the potential costs and benefits of different options^{2–5}. Here we use a large-scale epidemic simulation^{6,7} to examine intervention options should initial containment^{6,7} of a novel influenza outbreak fail, using Great Britain and the United States as examples. We find that border restrictions and/or internal travel restrictions are unlikely to delay spread by more than 2–3 weeks unless more than 99% effective. School closure during the peak of a pandemic can reduce peak attack rates by up to 40%, but has little impact on overall attack rates, whereas case isolation or household quarantine could have a significant impact, if feasible. Treatment of clinical cases can reduce transmission, but only if antivirals are given within a day of symptoms starting. Given enough drugs for 50% of the population, household-based prophylaxis coupled with reactive school closure could reduce clinical attack rates by 40–50%. More widespread prophylaxis would be even more logistically challenging but might reduce attack rates by over 75%. Vaccine stockpiled in advance of a pandemic could significantly reduce attack rates even if of low efficacy. Estimates of policy effectiveness will change if the characteristics of a future pandemic strain differ substantially from those seen in past pandemics.

We parameterize an individual-based simulation model of pandemic influenza transmission⁶ for Great Britain and the United States using high-resolution population density data⁸ and data on travel patterns (see Supplementary Information). We extend the model by incorporating realistic seeding of infection (via international travel) in the modelled countries, and by explicitly modelling air travel within the United States (air travel being relatively insignificant in Great Britain due to its much smaller size).

The model represents transmission in households, schools and workplaces, and the wider community. For the United States, best estimates are that 30% of transmission occurs within the household, and 70% outside the household (see Supplementary Information). Of the latter 70%, we assume that 33% occurs in the general community and 37% in schools and workplaces. To reproduce the higher attack rates seen in children in past pandemics, per-capita contact rates in schools were assumed to be double that in workplaces⁶. These assumptions affect estimates of the impacts of certain targeted control policies; for example, if school/workplace transmission were assumed to account for 50–60% of transmission, then policies such as school closure and socially targeted prophylaxis would be more effective (see Supplementary Information).

Acquiring more quantitative data on transmission in different social contexts should therefore be a priority.

We estimated the reproduction number⁹ for pandemic influenza, R_0 , to have a value of 1.7–2.0 for the first wave of the 1918 pandemic, as determined from city-level mortality data (see Supplementary Information). In 1957, epidemic growth rates were less, with UK national data giving R_0 values of 1.5–1.7 (see Supplementary Information). Inter-pandemic data give a value of $R_0 \approx 1.7$ (see Supplementary Information). We therefore examine values of R_0 in the range 1.4 to 2, particularly focusing on how conclusions differ for ‘moderate’ ($R_0 = 1.7$) and ‘high’ ($R_0 = 2.0$) transmission scenarios. Because the natural history of infection for human cases of avian H5N1 infection have to date been much more extended (and severe) than normal human influenza^{10,11}, we also examine sensitivity to assumptions about the duration of infectiousness. We do not assume any spontaneous change in the behaviour of uninfected individuals as the pandemic progresses, but note that behavioural changes that increased social distance together with some school and workplace closure occurred in past pandemics¹² (see Supplementary Information) and might be likely to occur in a future pandemic even if not part of official policy. Data on respiratory infection incidence in Hong Kong during the severe acute respiratory syndrome (SARS) epidemic supports this view¹³. Such spontaneous changes in population behaviour might more easily reduce peak daily case incidence than overall cumulative attack rates (see Supplementary Information). The peak incidence of the high transmissibility scenario examined here is therefore probably a worst case.

We assume that 50% (see Supplementary Information for sensitivity analysis) of those infected are ill enough to be classified as clinical cases (that is, those requiring and seeking medical care), consistent with known patterns for seasonal influenza. In reality, a spectrum of disease severity will be seen in a pandemic. Given the lack of data on the virulence of the next pandemic strain, the impact of antivirals on mortality, and the effects of improvements in general medical care, we examine the impact of interventions on clinical attack rates rather than mortality.

Figure 1 shows the expected pattern of spread of an influenza pandemic in Great Britain and the United States, for moderate and high transmissibility scenarios (Fig. 1a, b). The epidemic peaks some 50–65 days after the first case in the country for Great Britain, and some 60–80 days after the first case for the United States. Peak daily case incidence is slightly higher in Great Britain than the United States due to the smaller spatial scale of Great Britain and hence the more synchronized regional outbreaks. But peak local incidence in the United States is comparable with Great Britain, and cumulative incidence over the epidemic is identical for both countries. For the moderate and high transmissibility scenarios, 55% and 68% of the population are infected, respectively, giving cumulative clinical attack rates of 28% and 34% (namely, 84 million or 100 million

¹Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PG, UK. ²Department of International Health, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, Maryland 21205, USA. ³RTI International Inc., P.O. 12194, 3040 Cornwallis Rd, Research Triangle Park, North Carolina 27709, USA.

sick individuals in the United States). Absenteeism from work is, in reality, a complex function of case incidence, but can be approximated by case incidence multiplied by total numbers of work days lost per case. Assuming 7 work days lost per clinically ill case, nationwide absenteeism would peak at 10% in Great Britain and at 9% in the United States for the moderate transmissibility scenario. Local absenteeism can be substantially higher (each national epidemic is composed of slightly desynchronized local epidemics).

The dependence of the timing of the epidemic peak and the clinical attack rate on R_0 is shown in Fig. 1c. There is considerable stochastic variation in the timing of the first few cases, associated with the randomness of the disease importation process (Fig. 1d). But peak timing is much less variable. The rate at which infected persons enter a country from overseas will affect the likely rate of spread within a country. In the absence of a global simulation of pandemic spread, we modelled the global epidemic with a simple compartmental transmission model, and used international travel data to estimate the numbers of infected persons expected to arrive in either the United States or Great Britain per day (see Supplementary Information). The global model used was non-spatial, so it was not possible to examine how the country of origin of a pandemic would affect how quickly and in which order the United States and Great Britain would be infected. But use of even a non-spatial transmission model is preferable to simply assuming a fixed starting number of infected persons or a constant importation rate through time, as the expected exponential increase in imported infections over time significantly speeds spatial spread within the United States and Great Britain compared with assuming a single point seeding event.

Little spatial structure is seen in the pattern of pandemic establishment and spread in Great Britain (see Supplementary Videos) due to its relatively small size and frequent long-range travel. In the United States, more structure is apparent: early spread is focal around seed infections (typically in urban centres) imported from overseas, but rapidly becomes almost homogeneously distributed across the whole population (Fig. 1e and Supplementary Videos).

After the identification of a new human-to-human transmissible influenza A virus strain anywhere in the world, attempts to prevent spread to unaffected countries are likely, and international travel may be restricted^{14–16}. Here we examine the impact of the border controls

imposed by the United States or Great Britain to reduce numbers of inbound travellers (the results are very similar for both countries). Figure 2a shows how a 90%, 99% or 99.9% reduction in imported infections might delay the peak of the US pandemic by 1.5, 3, or 6 weeks, respectively (comparable delays would be expected for other Western countries given the similar mobility of their populations). To put these reductions into context, the 2003 SARS crisis resulted in an 80% reduction in travel to and from Hong Kong¹⁷. The magnitude of the impact of border controls is governed by the rate at which global infection prevalence increases. A tenfold reduction in numbers of visitors delays arrival of infection for approximately as long as it takes global prevalence to increase tenfold to compensate—12.5 days using the global model assumed here.

We next examined the effectiveness of travel restrictions within the United States at slowing national spread. If border controls are not in place, then external seeding from infected international travellers overwhelms the effect of within-country restrictions, only delaying the peak of the epidemic by less than 1 week. If combined with 99.9% effective border controls, blanket reductions in non-local travel achieve little in delaying the peak of the epidemic (Fig. 2b), but do reduce the peak attack rate substantially and spread the epidemic over a much longer time period (see Supplementary Information). Eliminating travel in and out of affected regions can delay spread by up to 2 weeks beyond the 6 weeks achieved by border restrictions alone. The delay achievable in Great Britain is half this. In the United States, closing airports for domestic travel also has little impact, due to the substantial volume of long distance travel by road. Internal restrictions must be highly effective to have much impact—although 90% effective internal restrictions can still have some effect, 75% effectiveness has almost none.

Once a new pandemic virus starts to be transmitted in a country, interventions must be targeted for maximum impact. Applying the type of intensive control strategies envisaged for containing a pandemic at source⁶ is impractical as infection will constantly be reseeded in a country by visitors (see Supplementary Information). Clinical cases are clearly then the first priority for any more-targeted approach, as prompt treatment with antivirals reduces clinical severity and infectiousness¹⁸ (see Supplementary Information). Our results indicate that only very rapid treatment can significantly

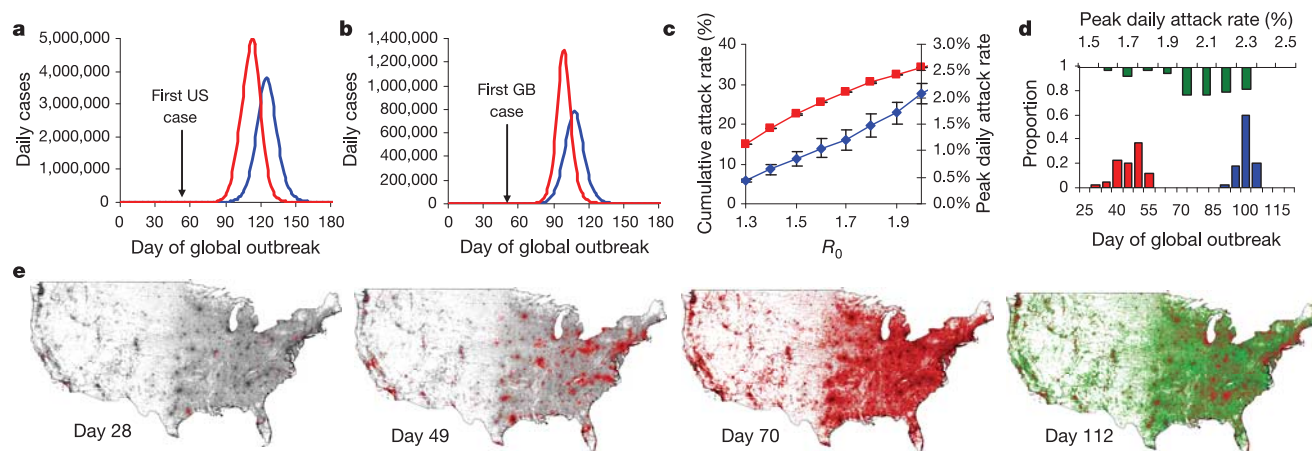


Figure 1 | Baseline pandemic dynamics. **a**, Clinical case incidence per day for the US pandemic (single realization shown) for high (red) and moderate (blue) transmissibility scenarios, assuming a generation time of 2.6 days, and that 50% of infected people are ill enough to be classified as clinical cases. Infection is seeded in the country as a function of the expected importation of infection from overseas derived from a simple global model of pandemic spread and available travel data (see Supplementary Information). Assumed population size of the United States was 300 million. Timing is shown both as days from the first case globally, and as days from the first case in the country. **b**, As **a**, but for Great Britain (modelled

population size 58.1 million). **c**, Cumulative (red) and peak daily (blue) clinical attack rates as a function of R_0 for Great Britain, averaged over 40 model realizations. Error bars show standard deviations. **d**, Histogram showing stochastic variability (across 40 model realizations) in timing of initial case (red), peak of epidemic (blue) and peak attack rate (green) for Great Britain ($R_0 = 2.0$). **e**, Snapshots of the extent of spread of the US pandemic (moderate transmissibility) at four time points. Greyscale indicates population density; red indicates areas with infective cases; and green indicates areas where the pandemic is over. See Supplementary Information for full parameter and model details.

reduce transmission (Fig. 2c, d), because cases are at their most infectious soon after symptoms develop (see Supplementary Information). For the high transmissibility scenario, same-day treatment of 90% of cases reduces cumulative attack rates from 34% to 29% and peak daily attack rates from 1.9% to 1.6%, with an antiviral stockpile sufficient to treat 25% of the population (the size many countries have ordered¹⁹) being adequate to implement the policy. If treatment is delayed by 1 day, the cumulative attack rate for the high transmissibility scenario increases to 32% (meaning that a 29% stockpile is needed), and the peak daily attack rate to 1.9%. The impact of treatment on the peak daily attack rate at the height of the epidemic is always greater than that on overall attack rates. Assuming that more than 50% of infections result in clinical illness requiring treatment would increase the required stockpile (see Supplementary Information). A real threat to the effectiveness of antiviral-based mitigation policies would be if resistant strains arose with transmissibility close to the wild-type level²⁰. Such strains have not yet been detected, but resistance monitoring during a pandemic will be essential.

Like treatment, rapid case isolation reduces the infectiousness of the targeted individual and can have a similar or greater impact (Fig. 2e)—reducing cumulative attack rates from 34% to 27% for $R_0 = 2.0$ if 90% of cases are isolated. Conservatively, we assume that clinical cases have a reduced contact rate in their school or workplace and with the wider community (see Supplementary Information) even in the absence of a policy of case isolation. Case isolation here is assumed to reduce all contact rates by a further uniform fraction,

including in the household—the policy is less effective if household contacts are not reduced.

Being a member of a household containing an influenza case is in fact the largest single risk factor for being infected oneself^{21,22}. Two policies targeting households are available (Fig. 3a–c): prophylaxis using antiviral drugs, and quarantine (requesting persons in households with infected cases to remain at home). Antiviral prophylaxis of household members is effective in reducing cumulative attack rates by at least one-third and peak attack rates by a half (Fig. 3a, b), but requires an antiviral stockpile large enough to treat 46% or 57% of the population for the moderate and high transmissibility scenarios, respectively. Household quarantine is also effective at reducing attack rates in the community (indeed, for low R_0 values, pandemic spread can be dramatically slowed), but only if compliance is high. However, given the expected increases in contact rates within the household which would result, a household quarantine policy might pose ethical dilemmas unless excellent infection control was implemented. A combined policy of household quarantine and household prophylaxis is arguably more feasible and could be highly effective.

School closure (Fig. 3d–f) causes a small reduction in cumulative attack rates, but a more substantial reduction in peak attack rates (of up to 40%). Such a reduction in peak incidence could mitigate stresses on healthcare systems and absenteeism in the critical workforce. Closure of 50% of workplaces can enhance the impact of school closure, but at higher economic cost (Fig. 3d–f). For a socially targeted prophylaxis policy, best use is made of drugs by targeting classmates or close work colleagues rather than the entire population of a school

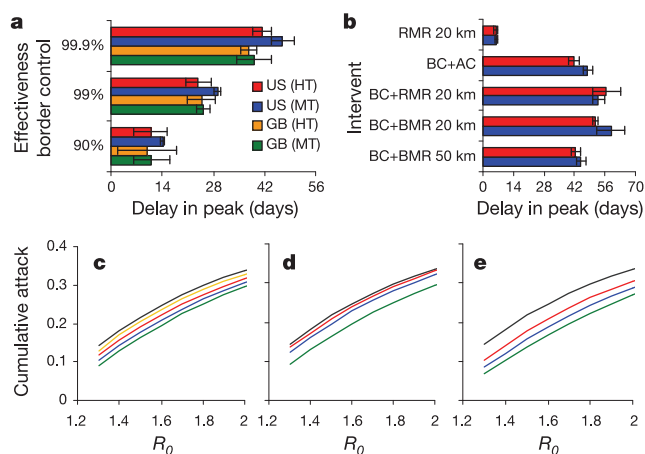


Figure 2 | Impact of travel restrictions and case-targeted policies. **a**, Delay caused by 90%, 99% and 99.9% reduction of imports of infection from day 30 of the global pandemic onwards on peak timing for GB and US pandemics for high (HT) and moderate (MT) transmissibility scenarios. Mean $\pm$ s.d. of 5–20 realizations is shown in both **a** and **b**. **b**, Delay in the peak of the US pandemic caused by internal travel restrictions for high (red) and moderate (blue) transmissibility. All policies are assumed to start after 50 cases have been diagnosed in the country. AC, all airports in the United States are closed to internal traffic; BC, border controls (external imports are reduced by 99.9%); BMR, blanket movement restrictions (journeys over 20 or 50 km from the home are eliminated); RMR, reactive movement restrictions (a 20-km exclusion zone is established around every diagnosed case, with overlapping zones being merged, and movement in and out of the exclusion zone is eliminated). **c**, Cumulative clinical attack rates for same day antiviral treatment policy, shown as a function of R_0 . From highest to lowest, the different curves represent 0%, 30%, 50%, 70% and 90% of cases treated. Results for Great Britain are shown (US results are identical). **d**, As **c**, but showing dependence on delay (in days) between symptom onset and treatment when 90% of cases are treated (curves, from highest to lowest, are for no treatment, 2 day, 1 day and 0 day delay). **e**, As **c**, but showing effect of same day case isolation causing a 90% reduction in contacts (from highest to lowest, curves are for 0%, 50%, 70% and 90% of cases isolated).

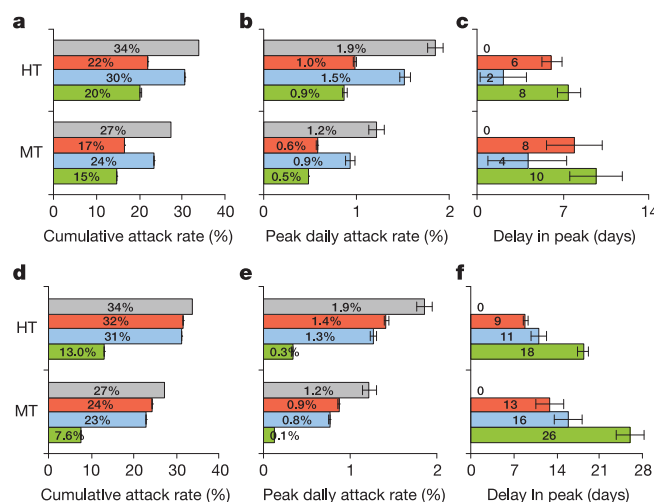


Figure 3 | Impact of household/socially targeted policies. Results shown for high (HT) and moderate (MT) transmissibility scenarios in the United States. **a–c**, Cumulative clinical attack rate (**a**), peak attack rate (**b**) and delay in peak achieved by policy (**c**) (as a percentage of total population size). Values in the absence of interventions are shown in grey. Three household policies are shown: red, treating 90% of clinical cases and applying prophylaxis to their households the day after symptoms start; blue, voluntary quarantine of households identified with a clinical case in the home for 14 days—50% of households are assumed to comply with the policy, and in these, external contact rates are reduced by 75% and within-household contact rates assumed to increase by 100%; green, combination of the previous two (red and blue) policies. **d–f**, As **a–c**, but for three school/workplace-targeted policies: red, reactive school closure (that is, closing 100% of schools (and 10% of workplaces) from the day after the first case in pupils or staff is detected until up to 3 weeks after the last case in the school)—contact rates in affected households are assumed to increase by 50% and community contact rates in absent staff/pupils by 25%; blue, as above (red) but assuming 50% of workplaces close; green, as household prophylaxis policy shown in **a–c** but adding prophylaxis of 90% of members of the same school class or work group as treated cases. Error bars show standard deviation of between 5 and 20 realizations.

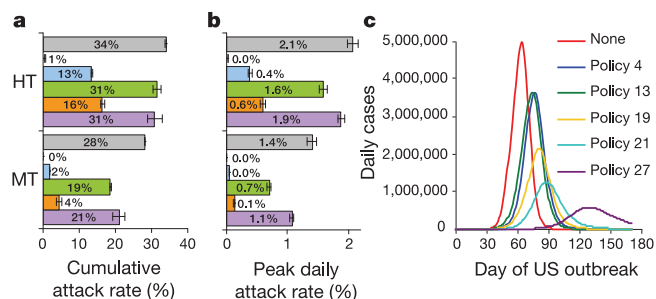


Figure 4 | Impact of vaccination and combination strategies. Results for the United States are shown. **a**, **b**, As Fig. 3a, b, but for a policy of mass vaccination assuming 1% of the population can be vaccinated per day from day 30 (red, second bar down), 60 (blue), 90 (green) after the first worldwide case, with 0–16 yr olds vaccinated first. Results for random (non-age prioritized) vaccination are also shown (orange), and for a policy of prioritizing those over 60 yr old (purple), both policies assuming that 1% of the population is vaccinated per day from day 60. Values in the absence of interventions are shown in grey. There is almost no effect of vaccination if started after 120 days. For comparison, the first US case is seen on day 47, and an untreated high transmissibility US pandemic peaks on day 113, on average. **c**, Example epidemic curves (for US high transmissibility scenario) for a selection of combination policies from Supplementary Table S11. Policies shown are: number 4 (50% household quarantine plus reactive school closure); 12 (as number 4 plus 50% next-day case treatment); 13 (90% case treatment plus reactive school closure); 19 (as number 13 plus household prophylaxis); 21 (as number 19 plus pre-vaccination of 20% of the population, prioritizing children); and 27 (as number 19 plus prophylaxis of 80% of school classmates and close work colleagues plus 99% effective border controls).

or workplace (see Supplementary Information). School/workplace prophylaxis could have a dramatic impact on attack rates (Fig. 3d–f), but this requires antiviral stockpiles of 72% or 102% of population size for the moderate and high transmissibility scenarios, respectively.

We now explore how rapidly vaccine needs to be deployed to have a significant impact. We assume a single dose vaccine giving protection in 2 weeks that reduces susceptibility by 70%, and infectiousness and probability of becoming a clinical case of those still getting infected by 30% and 50%, respectively. Vaccination at the rate of 1% of the population per day would need to begin within 2 months (Fig. 4a) of the initial global outbreak (or equivalently, at approximately the same time as the first cases are seen in the United States or Great Britain)—substantially faster than is possible using current vaccine technologies. A delay of 4 months from the start of the global pandemic would mean that the initial epidemic would be largely over by the time most of the modelled populations could be vaccinated. If two doses were required 1 month apart to achieve the same level of protection, then vaccination needs to start a month earlier still for the same impact. However a 20% stockpile of pre-prepared vaccine against the source avian virus could have a substantial impact on attack rates (Fig. 4c and Supplementary Table SI 1), even if its efficacy was less than a vaccine matched against the pandemic strain (we assume only a 30% reduction in susceptibility conferred). A staged vaccination programme has maximum effect on reducing transmission if children are vaccinated first, because school-age children have the highest transmission rates (Fig. 4a). Vaccinating the elderly first gives the lowest impact on transmission.

In reality, interventions will be applied in combination to reduce transmission in different social compartments simultaneously. We explore a representative sample of combination policies in the Supplementary Information, and show a subset of these in Fig. 4c. In addition to the impact of pre-vaccination, a number of other conclusions stand out. Household quarantine is potentially the most effective social distance measure, but only if compliance with the policy is good. Reactive school closure has limited impact on overall attack rates, but can enhance other policies. For the high

transmissibility scenario only policies using >99% border controls can delay spread by enough to enable pandemic vaccine to reduce attack rates substantially (vaccine is assumed to be produced from month 4 of the global epidemic).

The transmissibility of a future pandemic virus is uncertain, so we explored a number of scenarios here. It is also uncertain whether the generation time of a future pandemic strain would resemble that of typical human influenza, given the greater disease severity and more protracted course seen in patients infected with the avian H5N1 virus^{10,11}. Paradoxically, a virus that caused more severe and extended disease might be easier to control so long as R_0 was still comparable with previous pandemic strains. A further uncertainty is the proportion of transmission occurring in schools and workplaces. If the proportion is higher than we assume, the impact of school closure and school/workplace-targeted prophylaxis would increase. The impact of case isolation and household quarantine also depends on the assumptions made about the extent to which clinical cases and other members of the household reduce their contact rates while they are ill. Finally, the proportion of infections identified as clinical cases during the pandemic (here assumed to be 50%) can affect policy outcome significantly. See the Supplementary Information for discussion of these issues.

Lack of data prevent us from reliably modelling transmission in the important contexts of residential institutions (for example, care homes, prisons) and health care settings; detailed planning for use of antivirals, vaccines and infection control measures in such settings are needed, however. We do not present projections of the likely impact of personal protective measures (for example, face masks) on transmission, again due to a lack of data on effectiveness¹².

Although more detailed model validation and parameter estimation using data from past pandemics should be a priority for future research, it will be impossible to predict the exact characteristics of any future pandemic virus. If transmissibility is found to be more similar to the levels seen in 1968 or 1957 rather than in 1918, global spread will be slower, and all the non-travel-related control policies examined here will have substantially greater impact. If the duration of disease and shedding were to be extended compared with typical human influenza^{10,11}, then spread might be slower still, offering more potential for intervention. It will be imperative to collect the most detailed data on the clinical and epidemiological characteristics of a new virus and the impact of control measures early in the emergence of a pandemic, and to analyse those data in real time^{23,24} to allow interventions to be tuned to match the virus the world faces.

Received 25 January; accepted 11 April 2006.

Published online 26 April 2006.

1. World Health Organisation. WHO global influenza preparedness plan. (http://www.who.int/csr/resources/publications/influenza/GIP_2005_5Eweb.pdf) (2005).
2. Ferguson, N. M., Donnelly, C. A. & Anderson, R. M. Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. *Nature* **413**, 542–548 (2001).
3. Keeling, M. J. *et al.* Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* **294**, 813–817 (2001).
4. Riley, S. *et al.* Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions. *Science* **300**, 1961–1966 (2003).
5. Lipsitch, M. *et al.* Transmission dynamics and control of severe acute respiratory syndrome. *Science* **300**, 1966–1970 (2003).
6. Ferguson, N. M. *et al.* Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* **437**, 209–214 (2005).
7. Longini, I. M. Jr *et al.* Containing pandemic influenza at the source. *Science* **309**, 1083–1087 (2005).
8. Oakridge National Laboratory. Landscan global population data. (<http://www.ornl.gov/sci/gist/landscan/>) (2003).
9. Anderson, R. M. & May, R. M. *Infectious Diseases of Humans; Dynamics and Control* (Oxford Univ. Press, Oxford, 1991).
10. Chotpitayasunondh, T. *et al.* Human disease from influenza A (H5N1), Thailand, 2004. *Emerg. Infect. Dis.* **11**, 201–209 (2005).
11. The Writing Committee of the WHO Consultation on Human Influenza A/H5N1. Current concepts—Avian influenza A (H5N1) infection in humans. *N. Engl. J. Med.* **353**, 1374–1385 (2005).

12. World Health Organisation Writing Group. Non-pharmaceutical interventions for pandemic influenza, national and community measures. *Emerg. Infect. Dis.* **12**, 88–94 (2006).
13. Lo, J. Y. *et al.* Respiratory infections during SARS outbreak, Hong Kong, 2003. *Emerg. Infect. Dis.* **11**, 1738–1741 (2005).
14. Bell, D. M. Public health interventions and SARS spread, 2003. *Emerg. Infect. Dis.* **10**, 1900–1906 (2004).
15. Hufnagel, L., Brockmann, D. & Geisel, T. Forecast and control of epidemics in a globalized world. *Proc. Natl Acad. Sci. USA* **101**, 15124–15129 (2004).
16. World Health Organisation Writing Group. Non-pharmaceutical interventions for pandemic influenza, international measures. *Emerg. Infect. Dis.* **12**, 81–87 (2006).
17. Hong Kong International Airport. (<http://www.hongkongairport.com/eng/aboutus/statistics.html>) (2003).
18. Gani, R. *et al.* Potential impact of antiviral drug use during influenza pandemic. *Emerg. Infect. Dis.* **11**, 1355–1362 (2005).
19. Coombes, R. UK stocks up on antiviral drug to tackle flu outbreak. *Br. Med. J.* **330**, 495 (2005).
20. Ferguson, N. M., Mallett, S., Jackson, H., Roberts, N. & Ward, P. A population-dynamic model for evaluating the potential spread of drug-resistant influenza virus infections during community-based use of antivirals. *J. Antimicrob. Chemother.* **51**, 977–990 (2003).
21. Longini, I. M. J., Koopman, J. S., Monto, A. S. & Fox, J. P. Estimating household and community transmission parameters for influenza. *Am. J. Epidemiol.* **115**, 736–751 (1982).
22. Cauchemez, S., Carrat, F., Viboud, C., Valleron, A. J. & Boelle, P. Y. A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Stat. Med.* **23**, 3469–3487 (2004).
23. Ferguson, N. M. *et al.* Planning for smallpox outbreaks. *Nature* **425**, 681–685 (2003).
24. Ferguson, N. M., Fraser, C., Donnelly, C. A., Ghani, A. C. & Anderson, R. M.

Public health risk from the avian H5N1 influenza epidemic. *Science* **304**, 968–969 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the National Institute of General Medical Sciences MIDAS Program (N.M.F., D.A.T.C. and D.S.B.), the Medical Research Council (N.M.F.), the Royal Society (N.M.F. and C.F.) and the Howard Hughes Medical Institute (N.M.F.) for research funding. We thank B. Schwartz, R. Robinson, B. Gellin, D. Harper, J. Edmunds, P. Grove, R. Hatchett and members of the MIDAS consortium for useful discussions. We also thank the National Center for Supercomputing Applications (NCSA) and the MIDAS informatics group for computational resources and technical advice.

Author Contributions N.M.F. designed, implemented and ran the model, integrated the demographic and disease data sets used, and drafted and revised the text. All other authors edited or commented on the text. D.C. identified, collated and processed some demographic and travel data sets used and provided input on model assumptions. C.F. performed analytical modelling that aided the verification of the simulation and gave suggestions on model parameterization. J.C.C. assisted with the collation of demographic and travel data sets. P.C.C. assisted with provision of high performance computing facilities and technical advice. D.S.B. provided input into model design and assumptions, advised on the presentation of results and assisted with data collection.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.M.F. (neil.ferguson@imperial.ac.uk).

Endothelial tubes assemble from intracellular vacuoles *in vivo*

Makoto Kamei¹, W. Brian Saunders³, Kayla J. Bayless³, Louis Dye², George E. Davis³ & Brant M. Weinstein¹

The formation of epithelial tubes is crucial for the proper development of many different tissues and organs, and occurs by means of a variety of different mechanisms¹. Morphogenesis of seamless, properly patterned endothelial tubes is essential for the development of a functional vertebrate circulatory system, but the mechanism of vascular lumenization *in vivo* remains unclear. Evidence dating back more than 100 years has hinted at an important function for endothelial vacuoles in lumen formation². More than 25 years ago, in some of the first endothelial cell culture experiments *in vitro*, Folkman and Haudenschield described “longitudinal vacuoles” that “appeared to be extruded and connected from one cell to the next”^{3,4}, observations confirmed and extended by later studies *in vitro* showing that intracellular vacuoles arise from integrin-dependent and cdc42/Rac1-dependent pinocytic events downstream of integrin–extracellular-matrix signalling interactions^{5–10}. Despite compelling data supporting a model for the assembly of endothelial tubes *in vitro* through the formation and fusion of vacuoles, conclusive evidence *in vivo* has been lacking, primarily because of difficulties associated with imaging the dynamics of subcellular endothelial vacuoles deep within living animals. Here we use high-resolution time-lapse two-photon imaging of transgenic zebrafish to examine how endothelial tubes assemble *in vivo*, comparing our results with time-lapse imaging of human endothelial-cell tube formation in three-dimensional collagen matrices *in vitro*. Our results provide strong support for a model in which the formation and intracellular and intercellular fusion of endothelial vacuoles drives vascular lumen formation.

We were able to visualize vacuolar structures in living human vascular endothelial cells (ECs) cultured *in vitro* in three-dimensional (3D) collagen gel matrices by using light microscopy, the uptake of carboxyrhodamine or the expression of transgenes preferentially localizing to these structures (Fig. 1a–e). Large intracellular vacuoles are clearly visible within human ECs in 3D culture during lumen formation (Fig. 1a) and are readily labelled by uptake of carboxyrhodamine from the medium (Fig. 1b), indicating that they are pinocytic vacuoles. As we have described previously^{5,8–10}, we could also visualize pinocytic vacuolar structures in living ECs by perivacuolar localization of enhanced green fluorescent protein (EGFP)–cdc42wt (in which ‘wt’ represents wild-type) fusion protein (Fig. 1c). The cdc42 protein is prenylated with a C₂₀ geranylgeranyl group for membrane localization¹¹, so to provide an independent method of highlighting endothelial vacuoles in living cells we added a farnesylation site membrane localization signal¹² (designated by the suffix ‘-F’) to EGFP or mRFP1 (monomeric red fluorescent protein¹³). When expressed in ECs cultured in 3D collagen gels, EGFP-F protein displays perivacuolar localization (Fig. 1d–f).

To permit observation of the dynamics of endothelial vacuoles

and their role in vascular lumen formation *in vivo*, we expressed EGFP–cdc42wt or mRFP1-F in zebrafish ECs. The optical clarity and accessibility of zebrafish embryos and larvae make this an excellent model for imaging the dynamics of blood vessel formation *in vivo*^{14,15}. We began by establishing that ubiquitous expression of either EGFP–cdc42wt or mRFP1-F does not cause developmental defects in zebrafish embryos (data not shown). We then prepared constructs to drive the expression of these genes under the control of the zebrafish *flil* promoter, which we have used previously to drive endothelial expression in transgenic zebrafish¹⁶. We injected *flil:EGFP–cdc42wt* DNA into zebrafish embryos and derived a germline transgenic line, *Tg(fli1:EGFP–cdc42wt)*^{y48}, expressing EGFP–cdc42wt in angioblasts and ECs throughout the animal (see Methods). The *Tg(fli1:EGFP–cdc42wt)*^{y48} animals grow and develop indistinguishably from non-transgenic siblings, and adults have normal viability and fecundity (data not shown). We did not observe

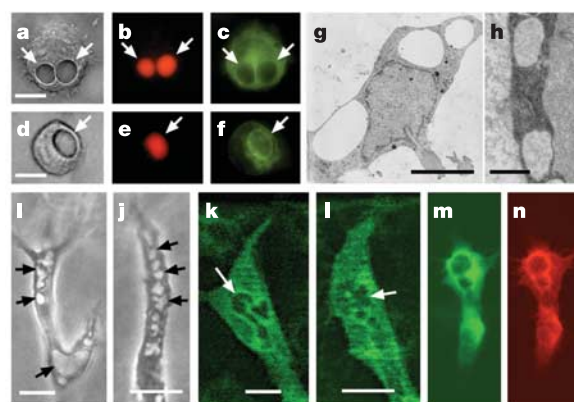


Figure 1 | Imaging of vacuoles in ECs *in vitro* and *in vivo*. **a–f**, Vacuoles (arrows) in living human umbilical-vein vascular ECs (HUVEC) cultured in 3D collagen gel matrices, revealed by light microscopy (**a**, **d**), epifluorescence of endocytosed carboxyrhodamine (**b**, **e**), and epifluorescence of either EGFP–cdc42wt (**c**) or EGFP-F (**f**) transgenes. **g**, **h**, Vacuolar structures revealed in transmission electron micrographs of fixed ECs in 3D EC cell cultures (**g**) and in the developing wall of the zebrafish dorsal aorta 24 h after fertilization (**h**). **i**, **j**, High-resolution light microscopic imaging of vacuoles (arrows) in EC in 3D culture *in vitro*. **k**, **l**, High-resolution multiphoton imaging of growing intersegmental vessels in *Tg(fli1:EGFP–cdc42wt)*^{y48} transgenic zebrafish, revealing similar vacuolar structures (arrows). **m**, **n**, Standard confocal imaging of doubly fluorescent ECs in *Tg(fli1:EGFP–cdc42wt)*^{y48} transgenic zebrafish injected with the *flil:mRFP1-F* construct, showing that the EGFP–cdc42wt (**m**) and mRFP1-F (**n**) proteins colocalize to intracellular vacuolar structures. Scale bars, 10 μ m (**a–g**), 1 μ m (**h**), 25 μ m (**i**, **k**), 15 μ m (**j**, **l**).

¹Laboratory of Molecular Genetics, and ²Microscopy and Imaging Core, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA. ³Department of Pathology, Texas A&M University System Health Science Center, 208 Reynolds Medical Building, College Station, Texas 77843-1114, USA.

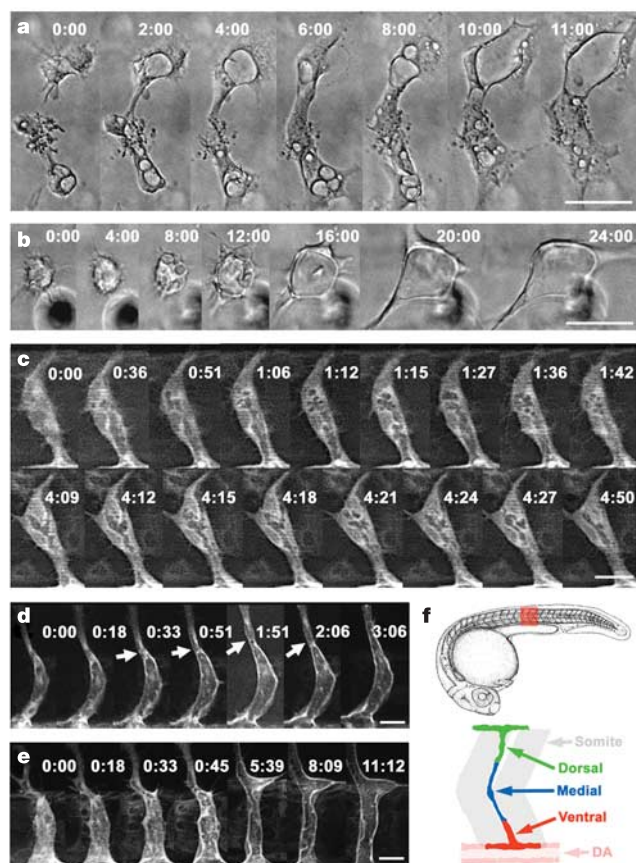


Figure 2 | Vacuoles are highly dynamic and fuse into large intracellular compartments. **a, b**, Time-lapse images of ECs in a 3D collagen gel, showing the formation of vacuoles and their fusion into larger intracellular compartments (**a**) and massive enlargement of the vacuolar compartment within a single EC (**b**). **c–e**, Two-photon time-lapse images of EGFP-positive ventral ECs in growing trunk intersegmental vessels in *Tg(fli1:EGFP-cdc42wt)^{y48}* transgenic animals, showing the emergence of vacuoles and their highly dynamic fusion into a large intracellular compartment (**c**), and relatively rapid (**d**) or more prolonged (**e**) formation of enlarged luminal spaces. Arrows in **d** indicate cell boundaries. **f**, Zebrafish intersegmental vessels are initially assembled from three cells arranged in a chain. The red box on the top drawing corresponds to the area depicted in **c–e**. See ref. 18 for additional information on zebrafish trunk vessel formation. DA, dorsal aorta. For each time-lapse sequence the time from the first frame of each successive image is shown in the format hours:minutes. Scale bars, 50 μ m (**a, b**), 20 μ m (**c–e**). See Supplementary Movies 1–5 for complete time-lapse sequences corresponding to **a–e**, respectively.

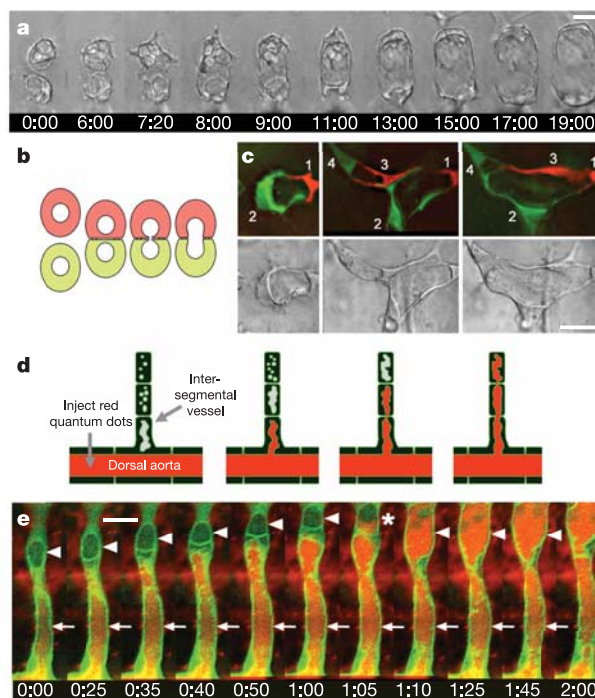


Figure 3 | Intracellular vacuolar compartments fuse to form intercellular luminal spaces. **a**, Time-lapse images of two EC cells in a 3D collagen-gel matrix showing fusion of their intracellular vacuoles to generate a single large luminal space. **b, c**, Fusion of intracellular vacuolar compartments preserves membrane topology and is not accompanied by cytoplasmic mixing. **b**, Differentially labelled ECs form junctions at points of cell-cell contact. Exocytosis of intercellular vacuoles into the cell-cell interface leads to formation of a common space bounded by both cells but sealed from the rest of the extracellular environment. **c**, Epifluorescence and light microscopy time-lapse images (top and bottom rows, respectively) of four HUVEC expressing either EGFP-cdc42 (green) or mRFP1-F (red) in a 3D collagen gel. The red and green fluorophores do not mix after intercellular vacuole fusion. **d, e**, Intercellular transfer of vacuolar contents in intersegmental vessels undergoing tubular morphogenesis *in vivo*. **d**, Red quantum dots injected into the circulation are transferred from the luminalized dorsal aorta into preformed vacuolar compartments within proximal intersegmental ECs and are subsequently transferred serially to more distal intersegmental ECs. **e**, Two-photon time-lapse imaging of trunk intersegmental vessel sprouts in living *Tg(fli1:EGFP-cdc42wt)^{y48}* transgenic animals injected intravascularly with 605-nm quantum dots. Quantum dots are transferred from the dorsal aorta to vacuoles in proximal (arrows) and then more distal (arrowheads) intersegmental ECs. In frame 1:05 the vacuole in the distal EC was captured in the process of filling with quantum dots from the proximal cell below (asterisk). For each time-lapse sequence the time from the first frame of each successive image is shown in the format hours:minutes. Scale bars, 25 μ m (**a, c**), 20 μ m (**e**). Supplementary Movies 7–10 show complete time-lapse sequences corresponding to **a, c** and **e**.

defects in vascular development in *Tg(fli1:EGFP-cdc42wt)^{y48}* embryos and larvae; vessels displayed normal kinetics of appearance, anatomy and morphology^{17,18}, as well as normal vessel pathfinding behaviour¹⁹ (data not shown).

We used two-photon imaging of the *Tg(fli1:EGFP-cdc42wt)^{y48}* line and/or confocal imaging of animals injected with the *fli1:mRFP1-F* construct to examine the formation and subsequent dynamics of vacuoles *in vivo*. We focused our initial efforts on the easily imaged, simple, and metamERICALLY repeating system of angiogenic trunk intersegmental vessels¹⁷. Previous studies have shown that each intersegmental vessel is initially formed from three ECs that emerge from the dorsal aorta and migrate as a chain along myotomal boundaries^{18–20}. Two-photon imaging of growing intersegmental vessels in living *Tg(fli1:EGFP-cdc42wt)^{y48}* embryos (Fig. 1k, l) revealed vacuolar structures similar to those observed in cultured human ECs during tube morphogenesis *in vitro* (Fig. 1i, j). Analogous subcellular structures were observed in confocal images of red fluorescent ECs in

wild-type animals injected with a *fli1:mRFP1-F* construct, indicating that the vacuolar structures we observed were not an artefact of EGFP-cdc42wt expression (data not shown). Colocalization of perivacuolar red and green fluorescence in ECs of *Tg(fli1:EGFP-cdc42wt)^{y48}* embryos injected with *fli1:mRFP1-F* showed that both fluorescent probes highlight the same subcellular structures (Fig. 1m, n). Endothelial vacuoles were also readily observed in transmission electron micrographs of newly formed blood vessels in wild-type (non-transgenic) zebrafish embryos, with an appearance similar to that of vacuoles in electron micrographs of ECs developing lumens in 3D collagen gels (Fig. 1g, h). Together, these results show that vacuoles are present in ECs *in vivo* and can be imaged in live animals with fluorescent probes that highlight these structures in ECs *in vitro*.

We used time-lapse imaging to examine the dynamics of endothelial vacuoles and their contribution to vascular lumen formation. Time-

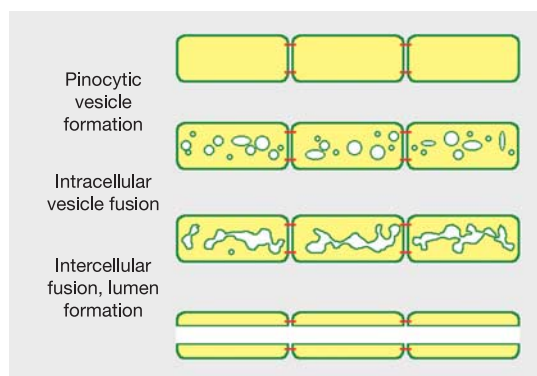


Figure 4 | A model for vascular lumen formation by intracellular and intercellular fusion of endothelial vacuoles. The diagram shows the mechanistic sequence of steps proposed to lead to intercellular lumen formation: intracellular vesicle formation, intracellular vesicle fusion, and finally intercellular merging of vacuolar compartments and lumen formation. These steps may not occur synchronously in all of the ECs in a lumenizing vessel but instead sequentially, as shown in Fig. 3d.

lapse imaging of cultured ECs reveals that endothelial vacuoles *in vitro* are highly dynamic, rapidly appearing and disappearing, fusing together and enlarging to create the luminal space (Fig. 2a, and Supplementary Movie 1). Over time the vacuoles enlarge and coalesce to form nascent luminal compartments encompassing most of the volume of the cell (Fig. 2b, and Supplementary Movie 2). High-resolution time-lapse two-photon imaging of lumenizing intersegmental vessels in *Tg(fli1:EGFP-cdc42wt)^{y48}* embryos with the use of a novel continuous-flow imaging chamber²¹ and a sensitive direct detector array revealed very similar vacuolar dynamics *in vivo* (Fig. 2c–e, and Supplementary Movies 3–6; vessels imaged in Fig. 2c–e are shown schematically in Fig. 2f). Vacuoles appear, disappear and fuse to form larger compartments on a timescale of minutes (Fig. 2c, and Supplementary Movie 3). At later stages, vacuoles merge into nascent luminal compartments filling most of the cell, as in ECs *in vitro* (Fig. 2d, e, and Supplementary Movies 4–6 provide examples of the range of behaviours observed in lumenizing vessels). These nascent luminal compartments remain extremely dynamic, with membrane protrusions continuously extending and retracting into the luminal space. Standard confocal imaging of red fluorescent ECs in wild-type embryos injected with the *fli1:mRFP1-F* construct revealed similar vacuolar dynamics (data not shown), indicating that our observations do not reflect abnormal consequences of EGFP-cdc42wt fusion protein expression.

We next investigated whether the enlarged vacuolar compartments that we observed within different ECs *in vitro* and *in vivo* underwent fusion to generate multicellular luminal spaces. Time-lapse microscopy of ECs in 3D collagen gels revealed ECs forming extensive cell–cell contacts followed by a merging of their intracellular vacuolar compartments (Fig. 3a, and Supplementary Movie 7). For a further examination of the nature of these fusion events *in vitro* we differentially labelled the cytoplasm of cultured ECs to determine whether intercellular vacuole fusion was accompanied by cytoplasmic fusion (Fig. 3b, c). We found that in all cases vacuole fusion occurred without cytoplasmic mixing (Fig. 3c, and Supplementary Movies 8 and 9), indicating that the formation of a common luminal space might occur by the exocytosis of intracellular vacuoles into junctional spaces between adjacent ECs (Fig. 3b). We performed additional experiments to determine whether fusion of preformed vacuolar compartments of separate ECs occurs *in vivo* and whether this is relevant to lumenization. We injected red quantum dots intravascularly to label the circulation of *Tg(fli1:EGFP-cdc42wt)^{y48}* embryos, and then examined whether the quantum dots were later transferred from the patent dorsal aorta to unlabelled vacuolar/nascent luminal compartments of ECs of the adjacent developing

intersegmental vessels (Fig. 3d, e). Serial transfer of the red fluorescent label could be observed from the dorsal aorta into previously unlabelled vacuolar compartments of the proximal intersegmental EC and then more distal ECs (Fig. 3e, and Supplementary Movie 10). These results indicate that preformed vacuolar compartments in adjacent ECs become linked to form a continuous luminal space in developing blood vessels *in vivo*.

Together, our results *in vitro* and *in vivo* support a proposed model in which the formation and intracellular and intercellular fusion of endothelial vacuoles drives vascular lumen formation (Fig. 4). Recent studies in both *Drosophila* and *Caenorhabditis elegans* indicate that intracellular lumen formation might be a common general mechanism for tube formation during development^{22–25}. We find no evidence for endothelial apoptosis during early vascular development, precluding models for vascular cavitation (death of the central cells in an endothelial cord to create a hollow space; V. Pham, personal communication). Although we have not yet performed a similar careful time-lapse analysis of larger-calibre blood vessels, lumen formation within larger aggregates or cords of ECs may occur through cord hollowing¹. This process would be mechanistically analogous to what we have observed for the intersegmental vessels—exocytosis of vacuoles into a common intercellular space bounded by multiple ECs (instead of just two cells) that are joined together by junctional contacts would lead to the formation of a common luminal space sealed off from the rest of the extracellular environment. It remains to be seen to what extent the molecular machinery involved in vascular lumen formation resembles that employed for other vacuolar/vesicular trafficking processes, such as vesicle trafficking at neuronal synapses^{26,27} or transvascular exchange and vascular permeability across endothelial monolayers^{28,29}.

METHODS

Zebrafish husbandry and generation of transgenic lines. Zebrafish husbandry and maintenance were performed as described³⁰. Molecular cloning and construction of vectors used for transgenesis in this study were conducted with standard methods. Transient and germline transgenic zebrafish lines used in this study were generated with methods described previously¹⁶. Additional details are given in Supplementary Methods.

Endothelial cell culture and 3D collagen vasculogenesis assays. Human umbilical vein EC culture and 3D collagen vasculogenesis assays were performed as described¹⁰. Additional details on cell culture and vectors and methods used for transfection and viral infection are given in Supplementary Methods.

Microscopy and imaging. Time-lapse imaging of zebrafish embryos was conducted as described^{14,21}. Intravascular injection of quantum dots was performed as described¹⁴. Additional details are given in Supplementary Methods. Time-lapse imaging of cultured human umbilical vein ECs was conducted with a temperature-controlled imaging chamber at 37°C with a continuous flow of 5% CO₂. Images were captured with a Nikon Eclipse TE2000-U inverted microscope. Additional details are given in Supplementary Methods.

Received 19 January; accepted 19 May 2006.

Published online 21 June 2006.

1. Lubarsky, B. & Krasnow, M. A. Tube morphogenesis: making and shaping biological tubes. *Cell* **112**, 19–28 (2003).
2. Downs, K. M. Florence Sabin and the mechanism of blood vessel lumenization during vasculogenesis. *Microcirculation* **10**, 5–25 (2003).
3. Folkman, J. & Haudenschild, C. Angiogenesis by capillary endothelial cells in culture. *Trans. Ophthalmol. Soc. U. K.* **100**, 346–353 (1980).
4. Folkman, J. & Haudenschild, C. Angiogenesis *in vitro*. *Nature* **288**, 551–556 (1980).
5. Bayless, K. J. & Davis, G. E. The Cdc42 and Rac1 GTPases are required for capillary lumen formation in three-dimensional extracellular matrices. *J. Cell Sci.* **115**, 1123–1136 (2002).
6. Bayless, K. J., Salazar, R. & Davis, G. E. RGD-dependent vacuolation and lumen formation observed during endothelial cell morphogenesis in three-dimensional fibrin matrices involves the $\alpha_v\beta_3$ and $\alpha_5\beta_1$ integrins. *Am. J. Pathol.* **156**, 1673–1683 (2000).
7. Davis, G. E. & Bayless, K. J. An integrin and Rho GTPase-dependent pinocytic vacuole mechanism controls capillary lumen formation in collagen and fibrin matrices. *Microcirculation* **10**, 27–44 (2003).
8. Davis, G. E., Bayless, K. J. & Mavila, A. Molecular basis of endothelial cell

- morphogenesis in three-dimensional extracellular matrices. *Anat. Rec.* **268**, 252–275 (2002).
9. Davis, G. E., Black, S. M. & Bayless, K. J. Capillary morphogenesis during human endothelial cell invasion of three-dimensional collagen matrices. *In Vitro Cell. Dev. Biol. Anim.* **36**, 513–519 (2000).
 10. Davis, G. E. & Camarillo, C. W. An $\alpha_2\beta_1$ integrin-dependent pinocytic mechanism involving intracellular vacuole formation and coalescence regulates capillary lumen and tube formation in three-dimensional collagen matrix. *Exp. Cell Res.* **224**, 39–51 (1996).
 11. Johnson, D. I. Cdc42: An essential Rho-type GTPase controlling eukaryotic cell polarity. *Microbiol. Mol. Biol. Rev.* **63**, 54–105 (1999).
 12. Hancock, J. F., Cadwallader, K. & Marshall, C. J. Methylation and proteolysis are essential for efficient membrane binding of prenylated p21K-ras(B). *EMBO J.* **10**, 641–646 (1991).
 13. Campbell, R. E. *et al.* A monomeric red fluorescent protein. *Proc. Natl Acad. Sci. USA* **99**, 7877–7882 (2002).
 14. Kamei, M., Isogai, S. & Weinstein, B. M. Imaging blood vessels in the zebrafish. *Methods Cell Biol.* **76**, 51–74 (2004).
 15. Weinstein, B. Vascular cell biology *in vivo*: a new piscine paradigm? *Trends Cell Biol.* **12**, 439–445 (2002).
 16. Lawson, N. D. & Weinstein, B. M. *In vivo* imaging of embryonic vascular development using transgenic zebrafish. *Dev. Biol.* **248**, 307–318 (2002).
 17. Isogai, S., Horiguchi, M. & Weinstein, B. M. The vascular anatomy of the developing zebrafish: an atlas of embryonic and early larval development. *Dev. Biol.* **230**, 278–301 (2001).
 18. Isogai, S., Lawson, N. D., Torrealday, S., Horiguchi, M. & Weinstein, B. M. Angiogenic network formation in the developing vertebrate trunk. *Development* **130**, 5281–5290 (2003).
 19. Torres-Vazquez, J. *et al.* Semaphorin-plexin signaling guides patterning of the developing vasculature. *Dev. Cell* **7**, 117–123 (2004).
 20. Childs, S., Chen, J. N., Garrity, D. M. & Fishman, M. C. Patterning of angiogenesis in the zebrafish embryo. *Development* **129**, 973–982 (2002).
 21. Kamei, M. & Weinstein, B. M. Long-term time-lapse fluorescence imaging of developing zebrafish. *Zebrafish* **2**, 113–123 (2005).
 22. Berry, K. L., Bulow, H. E., Hall, D. H. & Hobert, O. A. C. *C. elegans* CLIC-like protein required for intracellular tube formation and maintenance. *Science* **302**, 2134–2137 (2003).
 23. Buechner, M. Tubes and the single *C. elegans* excretory cell. *Trends Cell Biol.* **12**, 479–484 (2002).
 24. Manning, G. & Krasnow, M. in *The Development of Drosophila melanogaster* (eds Martinez-Arias, A. & Bate, M.) 609–685 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1993).
 25. Paul, S. M. & Beitel, G. J. Developmental biology. Tubulogenesis CLICs into place. *Science* **302**, 2077–2078 (2003).
 26. Rizzoli, S. O. & Betz, W. J. Synaptic vesicle pools. *Nature Rev. Neurosci.* **6**, 57–69 (2005).
 27. Sudhof, T. C. The synaptic vesicle cycle. *Annu. Rev. Neurosci.* **27**, 509–547 (2004).
 28. Carver, L. A. & Schnitzer, J. E. Caveolae: mining little caves for new cancer targets. *Nature Rev. Cancer* **3**, 571–581 (2003).
 29. Schnitzer, J. E. Caveolae: from basic trafficking mechanisms to targeting transcytosis for tissue-specific drug and gene delivery *in vivo*. *Adv. Drug Deliv. Rev.* **49**, 265–280 (2001).
 30. Westerfield, M. *The Zebrafish Book* (Univ. Oregon Press, Eugene, Oregon, 1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Faske for technical assistance, G. Martin for help in constructing the mRFP1-expressing human ECs, K. Tanegashima for assistance with western blot analysis, R. Tsien for providing the mRFP1 vector, and I. B. Dawid for critical reading of this manuscript. This work was supported in part by a grant from the NIH to G.E.D. B.M.W. is supported by the intramural program of the NICHD.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.M.W. (bw96w@nih.gov).

Electrical signals control wound healing through phosphatidylinositol-3-OH kinase- γ and PTEN

Min Zhao¹, Bing Song¹, Jin Pu¹, Teiji Wada², Brian Reid¹, Guangping Tai¹, Fei Wang^{3†}, Aihua Guo¹, Petr Walczysko¹, Yu Gu¹, Takehiko Sasaki⁴, Akira Suzuki⁵, John V. Forrester¹, Henry R. Bourne³, Peter N. Devreotes⁶, Colin D. McCaig¹ & Josef M. Penninger²

Wound healing is essential for maintaining the integrity of multicellular organisms. In every species studied, disruption of an epithelial layer instantaneously generates endogenous electric fields, which have been proposed to be important in wound healing^{1–3}. The identity of signalling pathways that guide both cell migration to electric cues and electric-field-induced wound healing have not been elucidated at a genetic level. Here we show that electric fields, of a strength equal to those detected endogenously, direct cell migration during wound healing as a prime directional cue. Manipulation of endogenous wound electric fields affects wound healing *in vivo*. Electric stimulation triggers activation of Src and inositol-phospholipid signalling, which polarizes in the direction of cell migration. Notably, genetic disruption of phosphatidylinositol-3-OH kinase- γ (PI(3)K γ) decreases electric-field-induced signalling and abolishes directed movements of healing epithelium in response to electric signals. Deletion of the tumour suppressor phosphatase and tensin homolog (PTEN) enhances signalling and electrotactic responses. These data identify genes essential for electrical-signal-induced wound healing and show that PI(3)K γ and PTEN control electrotaxis.

Endogenous wound electric fields were determined first more than 150 yr ago by the German physiologist Emil Du-Bois Reymond⁴. Such electric fields are generated when the epithelial layer is cut and the lesion short-circuits the transepithelial potential difference^{1,5–9}. Using various techniques, we confirmed consistent and sustained outward electric currents at wounds in human skin and in rodent cornea and skin (Fig. 1a, b). A large outward current of $4 \mu\text{A cm}^{-2}$ was measured at the wound edges of rat cornea and human skin. This current gradually increased to $10 \mu\text{A cm}^{-2}$ and persisted at $4–8 \mu\text{A cm}^{-2}$. The direction and magnitude of the current was independent of wound size and the current vector (the flow of positive charge) was directed towards the wound centre (Fig. 1a, b).

To test directly the effects of the electric signal on cell movement in wound healing, cell migration was monitored in monolayer epithelial cultures. In control monolayer corneal epithelium without an applied electric field (default healing), cells moved into the wounds in a coordinated manner¹⁰. When an electric field was applied with a polarity that opposed the default healing direction, the movement of the epithelium followed the direction of the electric signal and the wound opened up (Fig. 1c; Supplementary Fig. 1a). Reversal of the electrical polarity closed the wound (Fig. 1c, 99–204 min; Supplementary Movie 1). An electric field of 12.5 mV mm^{-1} with the cathode in the wound resulted in a significant increase ($P = 0.046$) in the distance of cell movement into the wound. Increasing the field strength

increased the speed of epithelial movement into the wound and this migration rate reached a maximum at $\sim 100–200 \text{ mV mm}^{-1}$ (Supplementary Fig. 1b). This field strength is comparable to the strength of endogenous wound electric fields (that is, $42–100 \text{ mV mm}^{-1}$), measured experimentally in animals and in humans^{1,11}. In addition, neutrophils and dermal fibroblasts that are also crucial for wound healing showed evident voltage- and time-dependent electrotactic responses (Supplementary Figs 2 and 3, and Movies 2 and 3). Notably, electrical cues also guide migration of stratified epithelium and control healing rates in a cornea whole-organ culture model¹² (Supplementary Fig. 4 and Movie 4). Thus, electrical signals are predominant directional cues that guide and stimulate the migration of inflammatory cells, fibroblasts and epithelial cells in wound healing.

To test the role of the endogenous electric field in wound healing, we manipulated transepithelial ion transport in epithelial wounds of rat cornea^{7,13,14}. In temporal and spatial maps of endogenous electric currents, Cl^- and Na^+ are the main components of electric currents in rat corneal wounds⁷ (Fig. 1a). Invariably, enhancing the ion flow increased endogenous wound electric fields and wound healing. AgNO_3 , which increases Cl^- efflux and Na^+ influx in corneal epithelium, significantly amplified the transcorneal potential difference ($P = 1.30 \times 10^{-7}$) (Supplementary Fig. 5a) and endogenous wound electric field ($P = 0.042$) (Supplementary Fig. 5b), resulting in augmented corneal wound healing *in vivo* (Supplementary Fig. 5c, d). By contrast, furosemide, which inhibits Cl^- efflux, significantly decreased the transcorneal potential difference ($P = 6.72 \times 10^{-8}$) resulting in a decreased endogenous wound electric field ($P = 0.01$), and impaired corneal wound healing (Supplementary Fig. 5). These results, together with our data on electric-field-regulated directional nerve growth and cell proliferation^{7,13,14}, suggest that endogenous transcellular potentials at wounds have an important role not only *in vitro* but also *in vivo*. The specificities of the agents used to manipulate endogenous electrical fields are discussed in the Supplementary Information.

How are electric migration cues relayed into cellular responses? To answer this question, we analysed whether electric fields induce specific signalling cascades similar to those observed in chemotactic cell migration^{15–22}. Intriguingly, exposure of both keratinocytes and neutrophils to electric fields in serum-free medium induced rapid and sustained phosphorylation of extracellular-signal-regulated kinase (ERK), p38 mitogen-activated kinase (MAPK), Src and Akt on Ser 473 (Fig. 1d). Phosphorylation of the Janus kinase JAK1 remained unchanged, indicating that electric currents activate only defined signalling pathways. Next, we examined the distribution of

¹School of Medical Sciences and Department of Ophthalmology, University of Aberdeen, Aberdeen AB25 2ZD, UK. ²Institute of Molecular Biotechnology of the Austrian Academy of Sciences, Dr Bohr-Gasse 3, A-1030 Vienna, Austria. ³Department of Cellular and Molecular Pharmacology, University of California, San Francisco, California 94143, USA. ⁴Department of Pathology and Immunology, and ⁵Department of Molecular Biology, Akita University School of Medicine, 1-1-1 Hondo, 010-8543 Akita, Japan. ⁶Department of Cell Biology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [†]Present address: Department of Cell and Developmental Biology, University of Illinois, B107 Chemical and Life Sciences Laboratory, 601 S. Goodwin Avenue, Urbana, Illinois 61801, USA.

activated Src and Akt in electrotactic cells because such intracellular signals are polarized in chemotaxis by means of activation of PI(3)K signalling at the leading edge, and lateral and back inhibition by PTEN in *Dictyostelium* or Rho in neutrophils^{15–22}. To test whether, as in classical chemotaxis, electric-field-induced signalling is polarized, we stained cells to assess the distribution of activated Src kinase. In electrotactic keratinocytes, phosphorylated Src indeed polarized in the migration direction (Fig. 1e).

We also examined the dynamics of phosphatidylinositol-3,4,5-trisphosphate (PtdIns(3,4,5)P₃) distribution in migrating HL60 cells using the PH domain of Akt fused to green fluorescent protein (PHAkt–GFP). Again, PtdInsP₃ was polarized to the leading edge of differentiated HL60 cells migrating towards the cathode. Reversal of electric field polarity caused a rapid redistribution of PHAkt–GFP towards the direction of migration (Supplementary Fig. 6 and Movie 5). Cells treated with latrunculin still polarized PHAkt–GFP, suggesting that distribution of PtdInsP₃ to the leading edge in electrotaxis is independent of actin polymerization (Supplementary Fig. 7). These data show that electrotactic cues activate defined signalling pathways and induce polarization of PI(3)K and Src pathways to the leading edge of migratory cells.

To exclude a possible involvement of chemotactic effects in electric-field-directed cell migration, we used a set-up in which continual flow of culture medium perpendicular to the electrical vector was maintained throughout the experiment²³; the directionality of

electric-field-induced cell migration was again unaffected (Supplementary Fig. 8). Moreover, we used a mutant *Dictyostelium* strain that lacks the G β subunit and therefore does not have a chemotactic response²⁴. Intriguingly, this mutant β^- *Dictyostelium* strain still showed robust electrotactic responses (Supplementary Fig. 9 and Movie 6). Although the interplay between chemical and electrical gradients will be extremely relevant *in vivo*, these results indicate that electrical stimuli can act independently of local chemical gradients and chemokine sensing.

To provide genetic proof that these above signalling cascades are indeed important for electrotaxis, we examined the role of PI(3)K γ in cells where its gene (phosphatidylinositol-3-kinase, catalytic, gamma subunit (*Pik3cg*); hereafter termed *p110 γ*) was disrupted²⁵ (Supplementary Fig. 10). *p110 γ ^{-/-}* cells showed impaired activation of Akt and partially reduced phosphorylation of Src, p38 and ERK in response to electric fields (Fig. 1d). Notably, loss of PI(3)K γ markedly abrogated electrotactic migration of epithelial cells in single-cell migration assays and monolayer wound healing assays (Fig. 2; Supplementary Fig. 11 and Movies 7 and 8). In addition, electrotactic directionality was attenuated in primary cultures of neutrophils (Supplementary Fig. 12 and Movie 9) and dermal fibroblasts (Supplementary Fig. 13 and Movie 3) from *p110 γ ^{-/-}* mice, confirming the importance of PI(3)K γ signalling in electrotactic responses. To exclude the possibility that the electrotactic phenotypes in *p110 γ ^{-/-}* cells were secondary to developmental alterations in these mutant mice, we blocked PI(3)K activity with wortmannin. Pharmacological

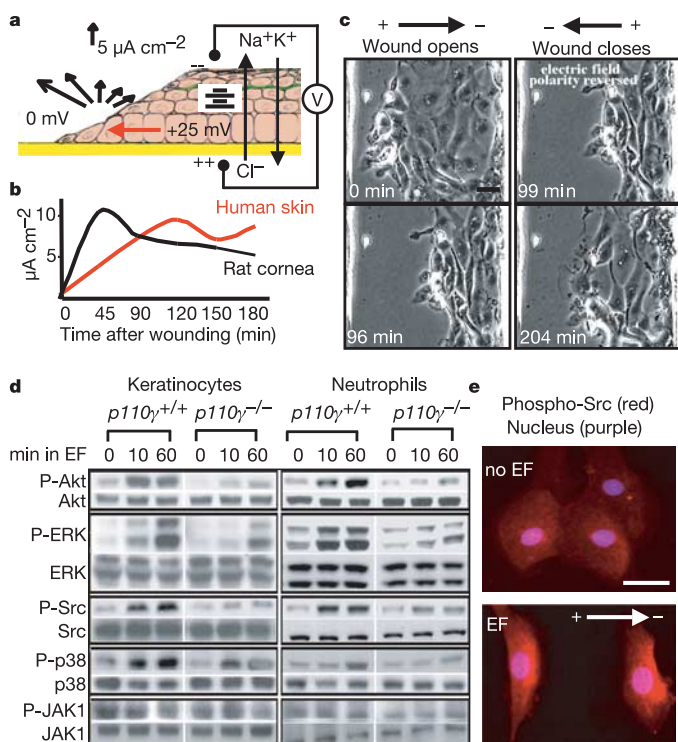


Figure 1 | Electrical signals direct cell migration in wound healing and activate selected signalling pathways. **a**, Wounding induces lateral electric fields directed towards the wound centre (red arrow), by collapsing the local transepithelial potential difference (V). Black arrows represent sizes and directions of currents. **b**, Directly measured currents increase over time in rat corneal and human skin wounds. **c**, **d**, An electric field (EF) directs migration of corneal epithelial cells in a monolayer model of wound healing (150 mV mm⁻¹; **c**) and activates Akt (Ser 473), Src (Tyr 416), ERK and p38 in primary cultures of mouse keratinocyte and mouse peritoneal neutrophils in serum-free medium (200 mV mm⁻¹; **d**). Disrupting *p110 γ* attenuates activation of these signalling pathways. Phosphorylated JAK1 and JAK1 are shown as controls. **e**, Phosphorylated Src kinase polarizes in the direction of cell migration in electrotactic mouse keratinocytes (150 mV mm⁻¹). Scale bar, 20 μ m.

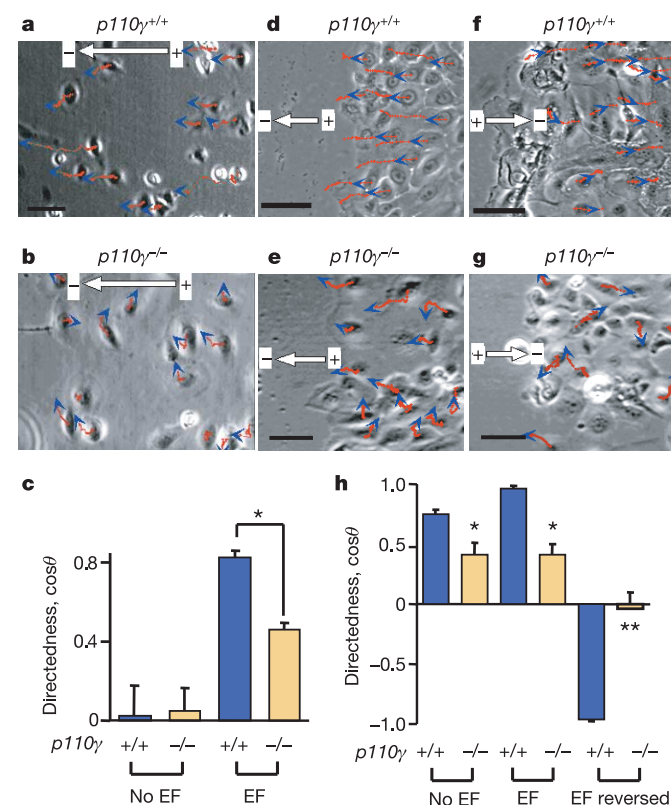


Figure 2 | Electrotaxis requires PI(3)K γ . **a–c**, Impaired electrotactic responses in nonconfluent *p110 γ ^{-/-}* keratinocytes. **d–h**, Attenuated electric field (EF)-directed migration of *p110 γ ^{-/-}* keratinocytes in monolayer wound healing assays. Wild-type (*p110 γ ^{+/+}*) cells are shown as controls. Electric fields that are applied with polarity opposite to the default healing direction direct the cells to move away from the wound (**f**, **g**). Red lines and blue arrows represent trajectories and direction of cell movement. Data (mean $\pm$ s.e.m.) were quantified from at least four independent experiments (**c**, **h**). **P* < 0.05; ***P* < 0.01, Student's *t*-test. Scale bars, 50 μ m, EF = 200 mV mm⁻¹. See also Supplementary Movies 7 (for **a–c**) and 8 (for **d–h**).

inhibition of PI(3)K again inhibited keratinocyte migration in response to electrical signals (Supplementary Figs 14 and 15, and Movies 10 and 11). These data show that PI(3)K γ controls electro-taxis and provide the first identification of a gene, *p110 γ* , that controls electric-field-induced cell migration.

To verify that PI(3)K γ regulates electrotaxis through PtdInsP₃ signalling, we investigated the effect of a tissue-specific deletion of the gene phosphatase and tensin homolog (*Pten*) in keratinocytes^{26,27} (Fig. 3a). The lipid phosphatase PTEN negatively regulates the PI(3)K/Akt pathway by downregulating the amount of PtdIns(3,4,5)P₃. Genetic disruption of *Pten* markedly enhanced electric-signal-induced ERK and Akt phosphorylation (Fig. 3b). Notably, *Pten* deletion enhanced electric-field-directed keratinocyte migration (Fig. 3c, d; Supplementary Movie 12). Moreover, in monolayer wound healing assays, loss of *Pten* enhanced electric-field-directed keratinocyte migration both into the wound and away from the wound with significantly higher directedness ($P = 0.022$ and $P = 0.017$ respectively) and increased migration rates ($P = 0.027$ and $P = 0.024$ respectively) when compared with *Pten*^{flax/flax} control keratinocytes (Fig. 3e–g; Supplementary Fig. 16 and Movie 13). These results indicate that PI(3)K γ and the tumour suppressor PTEN mediate directional sensing of cell migration in response to electric signals.

Because our data were obtained from cell-culture experiments, we

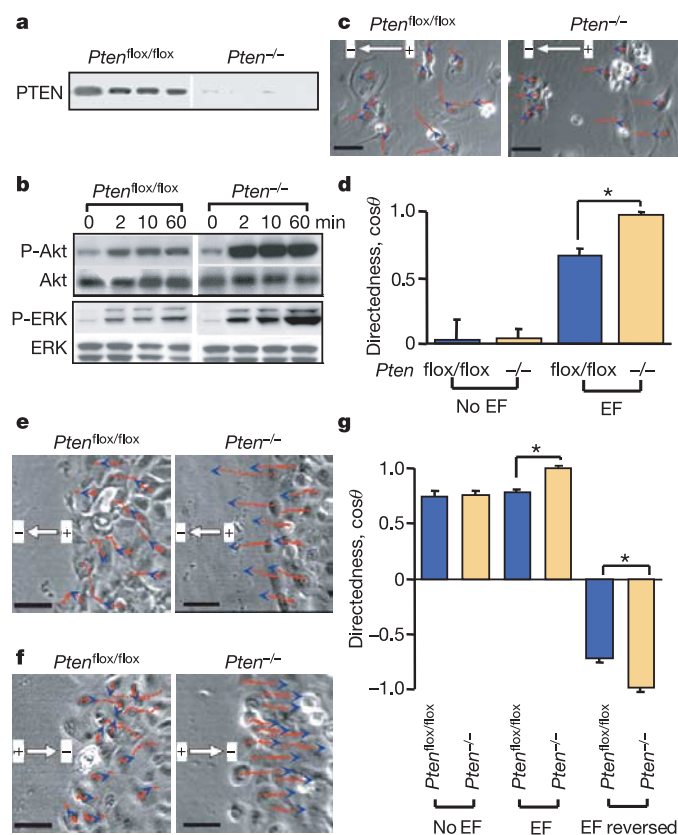


Figure 3 | The tumour suppressor PTEN negatively regulates electrotaxis. **a**, PTEN protein expression in keratinocytes. Four different cultures for each genotype are shown. **b**, Loss of PTEN expression in keratinocytes results in enhanced electric field (EF)-induced activation of Akt and ERK. **c**, **d**, Increased electrotactic migration of nonconfluent *Pten*-deficient keratinocytes. **e**–**g**, Loss of PTEN increases migration of keratinocytes in monolayer wound healing experiments in response to electric fields directed into (**e**, **g**) or away from (**f**, **g**) the wound. Red lines and blue arrows represent trajectories and direction of cell movement, respectively. Data are representative of at least four independent experiments with similar results. Quantification data are the mean $\pm$ s.e.m. (**d**, **g**). Scale bars, 50 μ m, EF = 200 mV mm⁻¹. * $P < 0.05$, Student's *t*-test. See also Supplementary Movies 12 (for **c**, **d**) and 13 (for **e**–**g**).

wanted to test whether this pathway is important for electrotactic wound healing after injury of a whole organ. We therefore tested whether PI(3)K γ signalling is important for the healing of stratified epithelium in cornea explant wounds¹². The direction and the magnitude of the wound electrical currents of cornea and skin from *p110 γ* ^{-/-} mice were similar to those from wild-type mice. In wild-type tissue, the wound edge of the corneal stratified epithelium moved into the wound, and this movement was significantly ($P = 0.027$) enhanced by an electric field applied with the cathode directed into the wound (Fig. 4a, b). Genetic disruption of *p110 γ* markedly impaired directed epithelial cell movement in response to electrical signals (Fig. 4c, d, g; Supplementary Movie 14). Application of an electric field with the polarity opposite to the default healing direction guided the stratified epithelial cells to migrate away from the wound, resulting in the wound opening up (Fig. 4e). Genetic disruption of *p110 γ* abolished this response (Fig. 4f, g; Supplementary Movie 14). Thus, PI(3)K γ expression is crucial for electrotaxis-regulated wound healing of a whole tissue.

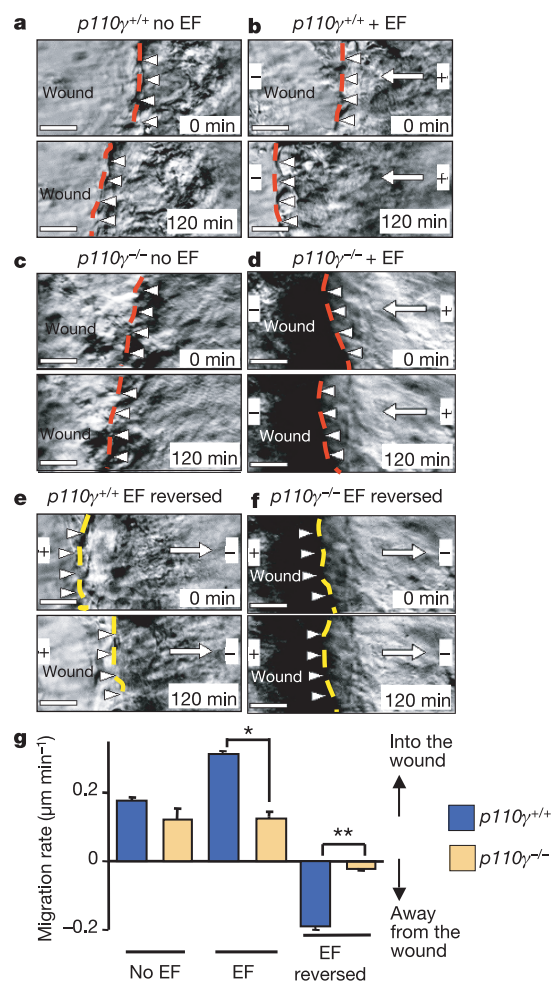


Figure 4 | PI(3)K γ is required for electrotactic cell movement in wound healing of stratified epithelium in *ex vivo* cornea cultures. **a**, Stratified corneal epithelium migrate *in situ* to heal a wound (towards the left). **b**, This wound healing response is significantly enhanced by an electric field with the cathode at the wound. **c**, **d**, Impaired electric-field-mediated wound healing in corneas isolated from *p110 γ* ^{-/-} mice. **e**, Electric fields applied with polarity opposite to the default healing direction direct the wound edge to migrate away from the wound. **f**, This response is impaired when *p110 γ* is disrupted. Results were confirmed in three or more independent experiments for each response. **g**, Quantification of the migration rates of the healing cornea epithelium from 3–7 experiments for a period of 120 min at each condition. Data are the mean $\pm$ s.e.m. EF = 150 mV mm⁻¹. Scale bars, 50 μ m. * $P < 0.05$, ** $P < 0.01$, Student's *t*-test. See also Supplementary Movie 14.

Because all cell types and intracellular organelles maintain transmembrane electrical potentials owing to asymmetric ion transport, wounding results in strong and directional ion flow after disruption of epithelial cell layers^{1,5–7}. To identify possible mediators that couple electric stimuli to intracellular responses, we tested the role of ion transporters in the electrotactic response. In particular, the Na⁺/H⁺ exchanger 1 (NHE1) has been implicated in directional cell migration²⁸. Two different types of NHE1 inhibitor, cariporide and EIPA²⁹, abrogated electric-field-induced Akt activation (Supplementary Fig. 17a) and decreased the directedness of cell migration in electric fields (Supplementary Fig. 17b). These results suggest that directional Na⁺/H⁺ transport by the NHE1 ion exchanger might relay the electric signal to PI(3)K activation with subsequent directional migration. In addition to Na⁺/H⁺ exchangers, it is likely that other ion channels such as Cl[−] channels (Supplementary Fig. 5) are also involved in electrotactic cell migration.

Although wound-induced electric currents have been known for more than 150 yr, the role of electrical signals in wound healing has long been discounted^{1–3,9}. Moreover, such responses have not been confirmed genetically. Using multiple model systems, we have shown that electric currents can act as directional cues in cell movement and wound healing. These cues activate signalling pathways similar to those reported for chemotaxis^{14–21}. Mechanistically, electric fields couple to directed cell migration through PI(3)Kγ and PTEN signalling. These experiments identify the first genes that modulate cell movements and wound healing in response to electrical currents.

METHODS

See Supplementary Information for full details.

Mutant mice, cell and tissue culture, and wound healing assays. *Pten*^{fllox} and *p110γ* mutant mice and the β[−] mutant *Dictyostelium* strain have been described^{24–27}. Primary cultures of keratinocytes from *Pten*^{fllox/fllox} mice were treated with adenovirus carrying GFP and Cre to delete the floxed *Pten* allele. In all experiments, littermate controls were used. Wound healing in corneal explant organ cultures and *in vivo* were done as described^{7,12}. All animal experiments were performed in accordance with institutional guidelines.

Electric fields. Endogenous wound electric currents were measured with a vibrating probe system⁵. DC electric fields of indicated strengths were applied in electrotactic chambers with modification for use in organ culture³⁰. Directedness was used as a parameter to define how cells migrate directionally in response to electric fields. Directedness values approaching one indicate migration directionally in the electric fields, whereas directedness values of or close to zero indicate random migration.

Western blot, immunofluorescence and time-lapse imaging. Primary cultures of keratinocytes and neutrophils, and wild-type fibroblasts were starved in serum-free medium before electric field stimulation. Cells were lysed and samples were probed with appropriate antibodies. For immunofluorescence microscopy, keratinocytes were exposed to electric fields, fixed, permeabilized for antibody labelling, and mounted in medium containing 4',6-diamidino-2-phenylindole dihydrochloride (DAPI). Dimethyl-sulphoxide-differentiated HL60 cells expressing PHAkt-GFP¹⁹ were exposed to an electric field. All time-lapse video images were recorded and analysed with a MetaMorph system.

Received 13 February; accepted 15 May 2006.

1. Barker, A. T., Jaffe, L. F. & Venable, J. W. Jr The glabrous epidermis of cavies contains a powerful battery. *Am. J. Physiol.* **242**, R358–R366 (1982).
2. Foulds, I. S. & Barker, A. T. Human skin battery potentials and their possible role in wound healing. *Br. J. Dermatol.* **109**, 515–522 (1983).
3. McCaig, C. D., Rajnicek, A. M., Song, B. & Zhao, M. Controlling cell behaviour electrically: current views and future potential. *Physiol. Rev.* **85**, 943–978 (2005).
4. DuBois-Reymond, E. Vorläufiger Abriss einer Untersuchung über den sogenannten Froschstrom und die electromotorischen Fische. *Ann. Phys. U. Chem.* **58**, 1–30 (1843).
5. Jaffe, L. F. & Nuccitelli, R. An ultrasensitive vibrating probe for measuring steady extracellular currents. *J. Cell Biol.* **63**, 614–628 (1974).
6. Borgens, R. B., Venable, J. W. Jr & Jaffe, L. F. Bioelectricity and regeneration: large currents leave the stumps of regenerating newt limbs. *Proc. Natl Acad. Sci. USA* **74**, 4528–4532 (1977).
7. Reid, B., Song, B., McCaig, C. D. & Zhao, M. Wound healing in rat cornea: the role of electric currents. *FASEB J.* **19**, 379–386 (2005).
8. Keesee, C. R., Wegener, J., Walker, S. R. & Giaever, I. Electrical wound-healing assay for cells *in vitro*. *Proc. Natl Acad. Sci. USA* **101**, 1554–1559 (2004).

9. Robinson, K. R. & Messerli, M. A. Left/right, up/down: the role of endogenous electrical fields as directional signals in development, repair and invasion. *BioEssays* **25**, 759–766 (2003).
10. Singer, A. J. & Clark, R. A. Cutaneous wound healing. *N. Engl. J. Med.* **341**, 738–746 (1999).
11. Chiang, M., Robinson, K. R. & Venable, J. W. Jr Electrical fields in the vicinity of epithelial wounds in the isolated bovine eye. *Exp. Eye Res.* **54**, 999–1003 (1992).
12. Zhao, M., Song, B., Pu, J., Forrester, J. V. & McCaig, C. D. Direct visualization of a stratified epithelium reveals that wounds heal by unified sliding of cell sheets. *FASEB J.* **17**, 397–406 (2003).
13. Song, B., Zhao, M., Forrester, J. & McCaig, C. Nerve regeneration and wound healing are stimulated and directed by an endogenous electrical field *in vivo*. *J. Cell Sci.* **117**, 4681–4690 (2004).
14. Song, B., Zhao, M., Forrester, J. V. & McCaig, C. D. Electrical cues regulate the orientation and frequency of cell division and the rate of wound healing *in vivo*. *Proc. Natl Acad. Sci. USA* **99**, 13577–13582 (2002).
15. Devreotes, P. & Janetopoulos, C. Eukaryotic chemotaxis: distinctions between directional sensing and polarization. *J. Biol. Chem.* **278**, 20445–20448 (2003).
16. Rickert, P., Weiner, O. D., Wang, F., Bourne, H. R. & Servant, G. Leukocytes navigate by compass: roles of PI3Kγ and its lipid products. *Trends Cell Biol.* **10**, 466–473 (2000).
17. Kimmel, A. R. & Parent, C. A. The signal to move: *D. discoideum* go orienteering. *Science* **300**, 1525–1527 (2003).
18. Firtel, R. A. & Chung, C. Y. The molecular genetics of chemotaxis: sensing and responding to chemoattractant gradients. *BioEssays* **22**, 603–615 (2000).
19. Servant, G. *et al.* Polarization of chemoattractant receptor signaling during neutrophil chemotaxis. *Science* **287**, 1037–1040 (2000).
20. Xu, J. *et al.* Divergent signals and cytoskeletal assemblies regulate self-organizing polarity in neutrophils. *Cell* **114**, 201–214 (2003).
21. Iijima, M. & Devreotes, P. Tumor suppressor PTEN mediates sensing of chemoattractant gradients. *Cell* **109**, 599–610 (2002).
22. Funamoto, S., Meili, R., Lee, S., Parry, L. & Firtel, R. A. Spatial and temporal regulation of 3-phosphoinositides by PI 3-kinase and PTEN mediates chemotaxis. *Cell* **109**, 611–623 (2002).
23. Erickson, C. A. & Nuccitelli, R. Embryonic fibroblast motility and orientation can be influenced by physiological electric fields. *J. Cell Biol.* **98**, 296–307 (1984).
24. Wu, L., Valkema, R., Van Haastert, P. J. & Devreotes, P. N. The G protein β subunit is essential for multiple responses to chemoattractants in *Dictyostelium*. *J. Cell Biol.* **129**, 1667–1675 (1995).
25. Sasaki, T. *et al.* Function of PI3Kγ in thymocyte development, T cell activation, and neutrophil migration. *Science* **287**, 1040–1046 (2000).
26. Suzuki, A. *et al.* T cell-specific loss of Pten leads to defects in central and peripheral tolerance. *Immunity* **14**, 523–534 (2001).
27. Suzuki, A., Sasaki, T., Mak, T. W. & Nakano, T. Functional analysis of the tumour suppressor gene PTEN in murine B cells and keratinocytes. *Biochem. Soc. Trans.* **32**, 362–365 (2004).
28. Denker, S. P. & Barber, D. L. Cell migration requires both ion translocation and cytoskeletal anchoring by the Na-H exchanger NHE1. *J. Cell. Biol.* **159**, 1087–1096 (2002).
29. Masereel, B., Pochet, L. & Laeckmann, D. An overview of inhibitors of Na⁺/H⁺ exchanger. *Eur. J. Med. Chem.* **38**, 547–554 (2003).
30. Zhao, M., Agius-Fernandez, A., Forrester, J. V. & McCaig, C. D. Orientation and directed migration of cultured corneal epithelial cells in small electric fields are serum dependent. *J. Cell Sci.* **109**, 1405–1414 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from the Wellcome Trust and Royal Society (to M.Z.), the Institute of Molecular Biotechnology of Austria of the Austrian Academy of Sciences, the Austria Ministry of Sciences, the Austrian National Bank, and a European Union Framework 6 excellence grant (all to J.M.P.). We thank all members of the Zhao, Bourne and Penninger laboratories for discussions and technical assistance.

Author Contributions M.Z. designed the experiments, took part in the cell migration and western blotting experiments, analysed the results and wrote the paper. J.M.P. designed the genetic analysis of the electric-field-induced signalling pathway, analysed the data and wrote the paper. B.S. did the *in vivo* experiments, most experiments with cells and tissues from transgenic mice. B.S. and J.P. did most of the cell migration and wound healing assays. T.W. performed the first signalling experiment and genotyping. B.R. performed the vibrating probe measurements. G.T., F.W. and P.W. did the experiments with HL60 cells. B.S., A.G. and Y.G. did the experiments on fibroblasts. P.N.D., A.S. and T.S. provided mouse and *D. discoideum* lines essential for the experiments. J.V.F., H.B. and C.D.M. helped with some of the experimental design, writing and analysis of the data. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.Z. (m.zhao@abdn.ac.uk) or J.M.P. (josef.penninger@imba.oew.ac.at).

IL-23 promotes tumour incidence and growth

John L. Langowski^{1*}, Xueqing Zhang^{1*}, Lingling Wu¹, Jeanine D. Mattson¹, Taiying Chen¹, Kathy Smith¹, Beth Basham¹, Terrill McClanahan¹, Robert A. Kastelein¹ & Martin Olt¹

Chronic inflammation has long been associated with increased incidence of malignancy and similarities in the regulatory mechanisms have been suggested for more than a century¹. Infiltration of innate immune cells, elevated activities of matrix metalloproteases and increased angiogenesis and vasculature density are a few examples of the similarities between chronic and tumour-associated inflammation². Conversely, the elimination of early malignant lesions by immune surveillance, which relies on the cytotoxic activity of tumour-infiltrating T cells or intra-epithelial lymphocytes, is thought to be rate-limiting for the risk to develop cancer³. Here we show a molecular connection between the rise in tumour-associated inflammation and a lack of tumour immune surveillance. Expression of the heterodimeric cytokine interleukin (IL)-23, but not of its close relative IL-12, is increased in human tumours. Expression of these cytokines antagonistically regulates local inflammatory responses in the tumour microenvironment and infiltration of intra-epithelial lymphocytes. Whereas IL-12 promotes infiltration of cytotoxic T cells, IL-23 promotes inflammatory responses such as upregulation of the matrix metalloprotease MMP9, and increases angiogenesis but reduces CD8 T-cell infiltration. Genetic deletion or antibody-mediated elimination of IL-23 leads to increased infiltration of cytotoxic T cells into the transformed tissue, rendering a protective effect against chemically induced carcinogenesis. Finally, transplanted tumours are growth-restricted in hosts depleted for IL-23 or in IL-23-receptor-deficient mice. Although many strategies for immune therapy of cancer attempt to stimulate an immune response against solid tumours, infiltration of effector cells into the tumour tissue often appears to be a critical hurdle^{4–7}. We show that IL-23 is an important molecular link between tumour-promoting pro-inflammatory processes and the failure of the adaptive immune surveillance to infiltrate tumours.

IL-12 and IL-23 are members of a small family of pro-inflammatory heterodimeric cytokines⁸. Both cytokines share a common p40 subunit that is covalently linked either to a p35 subunit to form IL-12 or to a p19 subunit to form IL-23 (ref. 9). The receptor for IL-12 is comprised of an IL-12R β 1 and IL-12R β 2 subunit, whereas the receptor for IL-23 is comprised of the IL-12R β 1 subunit and a novel component termed IL-23R¹⁰. Both cytokines are expressed predominantly by activated dendritic cells and phagocytic cells. Receptors for both cytokines are expressed on T cells, natural killer cells and natural killer T cells, but low levels of IL-23 receptor complexes are also found on monocytes, macrophages and dendritic-cell populations¹⁰. Despite these similarities, there is increasing evidence that IL-12 and IL-23 drive divergent immunological pathways. Whereas IL-12 leads to development of 'classical' interferon- γ -producing Th1 cells and enhances cytotoxic, antimicrobial and anti-tumour responses, IL-23 drives a pathway that leads to the generation of IL-17-producing CD4⁺ T cells^{11–13}. The induction of IL-23-derived

processes leads to recruitment of a range of inflammatory cells as well as T_HIL-17 T cells, and has been shown to be crucial to the pathogenesis of a number of immune-mediated inflammatory diseases¹⁴.

Given the combined evidence of a causal relationship between chronic inflammation and cancer as well as the pivotal role IL-23 plays in autoimmunity, we hypothesized that IL-23-mediated responses may be important in tumour promotion. To test this, we employed a comprehensive examination of messenger RNA expression within various human cancer panels. Surprisingly, IL-23p19 mRNA was found significantly upregulated in the overwhelming majority of carcinoma samples from various organ types when compared with their adjacent normal tissue (Fig. 1a, b). Significant mRNA upregulation was observed for both subunits of IL-23 (Fig. 1c), but not for the IL-12-specific subunit p35 (Supplementary Fig. S2). The expression of IL-17, a cytokine central to tumour-promoting angiogenesis¹⁵, was also found to be significantly elevated in human tumours, consistent with activation of IL-23-induced processes (Supplementary Fig. S2 and data not shown). Additionally, human colorectal tumours present a high degree of correlation between increased IL-23p19 expression and increased macrophage and neutrophil infiltration activation¹⁶ (Supplementary Fig. S3). Expression of IL-23p19 protein was localized within human tumour tissue but not in the surrounding stroma, and observed on infiltrating CD11c⁺ dendritic cells as a dominant source of the cytokine (Supplementary Fig. S4a, b). Supporting this, tumour-infiltrating macrophages (CD11b⁺) and dendritic cells (CD11c⁺) isolated from mouse syngeneic tumour models expressed preferentially high amounts of IL-23p19 mRNA rather than IL-12p35 when compared to those isolated from the spleen (Supplementary Fig. S4c, d).

To examine more directly the role of IL-12 and IL-23 in epithelial tumorigenesis, we employed a strategy of genetic deletion, analysing the susceptibility of mice deficient in either IL-12 or IL-23 (*p35*^{-/-} or *p19*^{-/-} mice) or mice deficient in both cytokines (*p40*^{-/-} mice) to tumour formation during chemical carcinogenesis¹⁷. Whereas *Il12p35*^{-/-} mice showed earlier appearance and developed significantly increased numbers of papillomas compared to C57BL/6 control mice (Fig. 2a, b), mice deficient in IL-23p19 were resistant to tumour induction. Surprisingly, *p40*^{-/-} mice, deficient in both cytokines, were resistant to tumour induction, emphasizing the importance of IL-23 for increased tumour incidence. Papillomas generated by this protocol persisted well over 100 days after cessation of TPA promotion (see Methods) and often progressed towards malignancy (Fig. 2b and data not shown); however, the low overall incidence of benign tumours in the absence of IL-23 prevents a statistically relevant comparison with WT mice in this regard.

To explore the role of IL-23 in tumour susceptibility, we examined the expression of inflammatory immune cell products in carcinogen-treated skin. IL-23 is known to modulate the homeostasis of

¹Schering-Plough BioPharma (formerly DNAX Research, Inc.), 901 California Avenue, Palo Alto, California 94304, USA.

*These authors contributed equally to this work.

neutrophil infiltration into tissues by inducing expression of IL-17 and G-CSF¹⁸—two cytokines that are also linked to tumour growth^{15,19}. We found IL-17 highly expressed in the hyperplastic skin of both wild-type C57BL/6 (WT) and *Il12p35*^{-/-} mice, but barely detectable in *Il23p19*^{-/-} and *Il12p40*^{-/-} mice. Similarly, G-CSF levels were decreased in *Il23p19*^{-/-} mice (Fig. 2c). The number of granulocytes (GR1⁺) and macrophages (CD11b⁺, F4/80⁺) in the dermis of *Il23p19*^{-/-} and *Il12p40*^{-/-} mice was also decreased compared to the skin of C57BL/6 and *Il12p35*^{-/-} mice (Fig. 2d, Supplementary Fig. S5a, b and data not shown). In murine cancer models, tumour development is dependent on MMP9 activity produced by inflammatory macrophages and mast cells²⁰. Additionally, MMP9 and IL-17 have been identified as both markers and mediators of tumour angiogenesis^{15,21}. The expression of activated MMP9 was elevated in carcinogen-treated skin of tumour-prone WT and *Il12p35*^{-/-} animals, but dramatically lower in skins of *Il23p19*^{-/-}

and *Il12p40*^{-/-} mice (Fig. 2e–h, Supplementary Fig. S5d, e). Angiogenic markers also followed this pattern, with significantly lower vessel density in IL-23p19-deficient animals (Fig. 2i–l, Supplementary Fig. S5c). Thus, the absence of IL-23 resulted in a significant reduction in numerous inflammatory factors and cell types essential for tumour promotion.

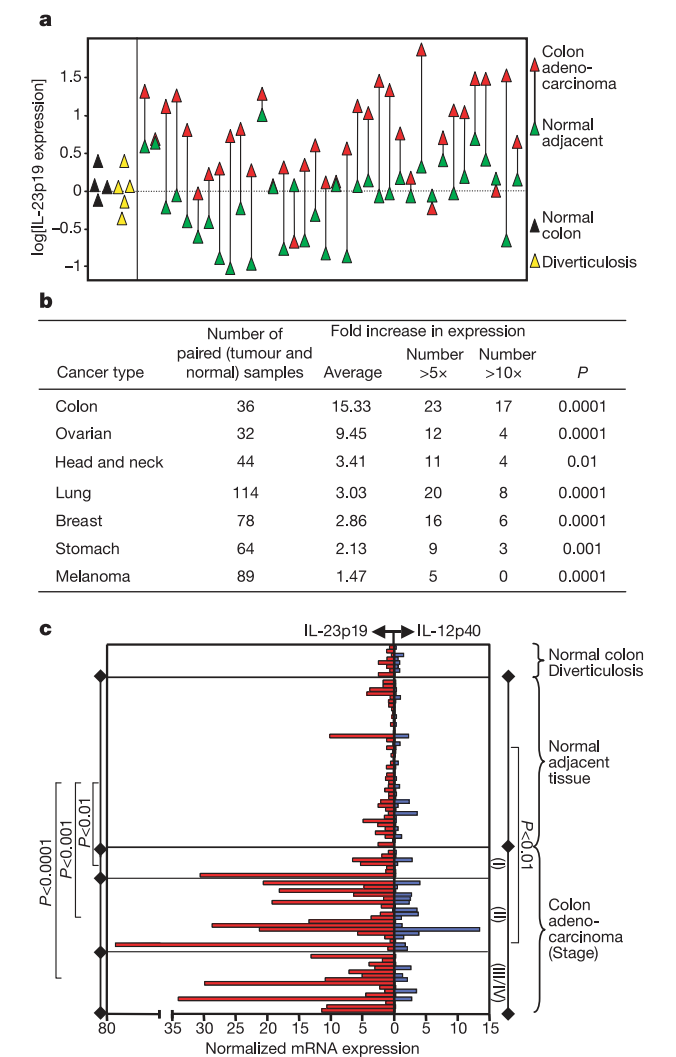
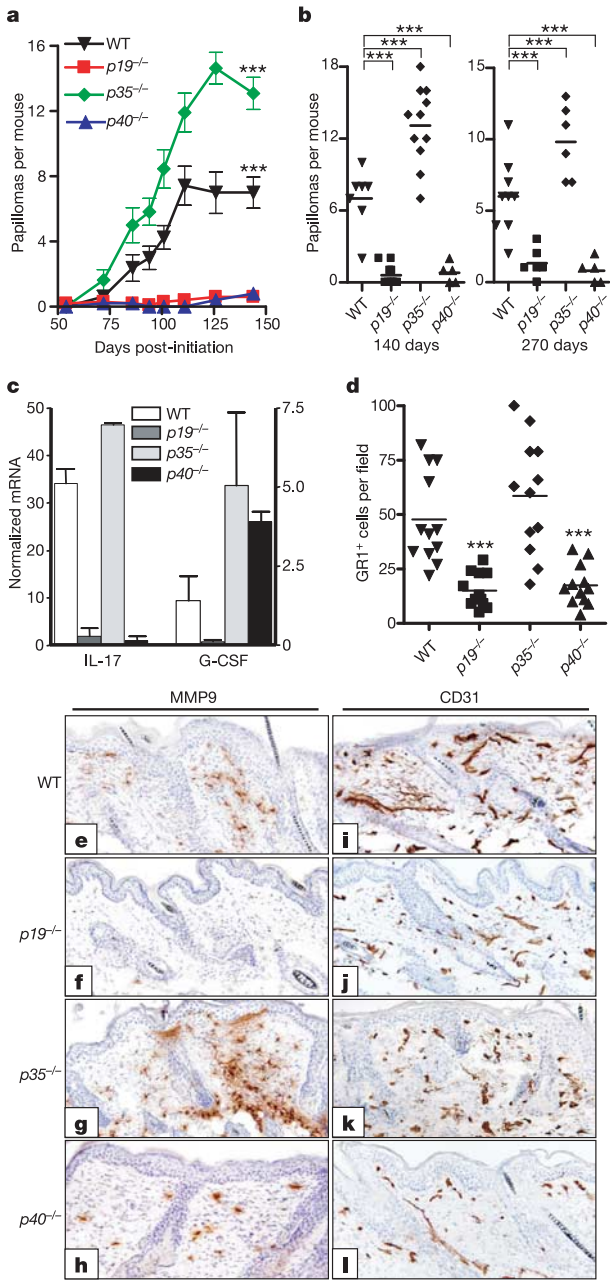


Figure 1 | Overexpression of IL-23 but not of IL-12 in human cancer. **a**, Quantitative mRNA expression of IL-23p19 in human colon tumour (red) compared to normal adjacent tissue of the same individual (green, connected by a line), or tissue from cancer-free individuals (black, yellow). **b**, Significant upregulation of IL-23p19 expression in various human cancers. The table shows number of individual paired cancerous and normal adjacent samples with average fold increase in expression, number of samples with five- and ten-fold increase, and *P*-value. **c**, Expression of both IL-23 subunits is increased in human cancer; each *y*-axis position signifies expression within an individual sample. Histological classification of samples as normal, normal adjacent and cancerous (with stage classification) is given.



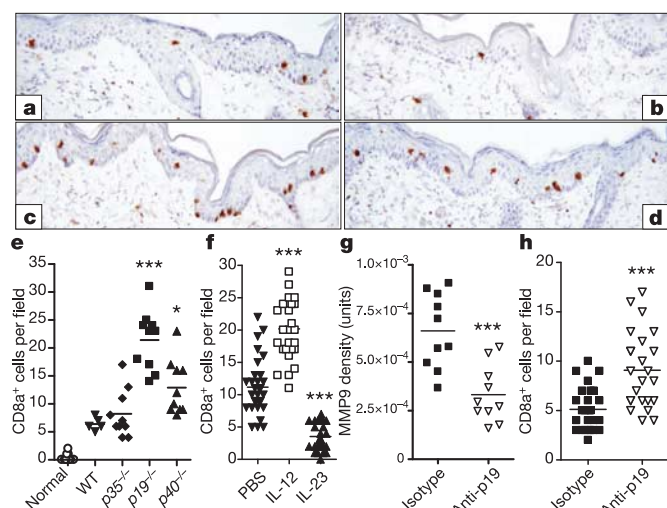


Figure 3 | IL-12 stimulates, but IL-23 decreases, immune surveillance by CD8 T cells. **a–d**, Increased immune surveillance by CD8 T cells in the epidermis of *Il23p19*^{-/-} mice (**c**) and *Il12p40*^{-/-} mice (**d**) treated with carcinogen for 140 days, compared to WT (**a**) and *p35*^{-/-} (**b**). Quantification of infiltrating CD8 T cells for each genotype is shown in **e**. **f**, Alterations in epithelial CD8 infiltration following treatment with exogenous IL-23 or IL-12 compared to control-treated animals. **g, h**, Systemic treatment with anti-IL-23p19 antibody shows marked decrease in MMP9 expression (**g**), but increase in CD8⁺ T cells (**h**), compared to control. Scatter plots show incidence or density with bars depicting the mean. Significance of results in figures is represented by: **P* < 0.05 and ****P* < 0.001; *n* = 3 per group.

Because the incidence of macroscopic tumours can be viewed as the result of a disturbed balance between tumour-promoting factors and tumour-cell-eliminating mechanisms, our observations raised the question of whether endogenous IL-23 would have a direct role in

immune surveillance against tumours. Immune surveillance and anti-tumour responses are vigorously promoted by IL-12, which induces proliferation and cytotoxic activity of natural killer cells and CD8⁺ T cells²². Surprisingly, the carcinogen-induced hyperplastic skin of *Il23p19*^{-/-} and *Il12p40*^{-/-} mice were infiltrated with a dramatically increased number of cytotoxic CD8⁺ T cells in both the dermis and the epidermis when compared to WT and *p35*^{-/-} mice (Fig. 3a–e). The expression of markers associated with the activity of cytotoxic T cells such as Fas ligand, perforin and granzymes were decreased in *Il12p35*^{-/-} animals but increased in *Il23p19*^{-/-} mice (Supplementary Fig. S6 and data not shown). Although deficiency of IL-12 or IL-23 does not alter the development of CD8 cells (ref. 13 and data not shown), it may alter the response necessary for tissue infiltration. To test this, we changed the microenvironment of the carcinogen-treated skin of C57BL/6 mice by serial intradermal injection of IL-12 and IL-23. While administration of IL-12 led to increased immune surveillance, as shown by the influx of greater numbers of CD8⁺ T cells in the epidermis, IL-23 injection caused a significant decrease of CD8 T cells in the hyperplastic skin (Fig. 3f, Supplementary Fig. S7a–c). Cytotoxic activity markers such as perforin, granzyme, Fas ligand and interferon-γ increased on IL-12 administration, whereas IL-23 greatly raised mRNA levels of IL-17 (Supplementary Fig. S7d, e). Systemic treatment with neutralizing antibodies specific to IL-23p19 led to a marked down-modulation of MMP9 expression and increased surveillance of CD8⁺ T cells within the carcinogen-treated skin (Fig. 3g–h, Supplementary Fig. S8). Taken together, these results provide evidence that IL-23 negatively influences both the number and activity of cytotoxic T cells within transformed tissue.

To further investigate the effects of IL-12 and IL-23 on immune surveillance, IL-12- and IL-23-cytokine-deficient mice were challenged intradermally with PDV squamous carcinoma cells¹⁷. In accordance with the dominant role of IL-12 in tumour immunity, deficiency of IL-12p35 or IL-12p40 dramatically increased tumour incidence, indicating the diminished ability of the host to reject the

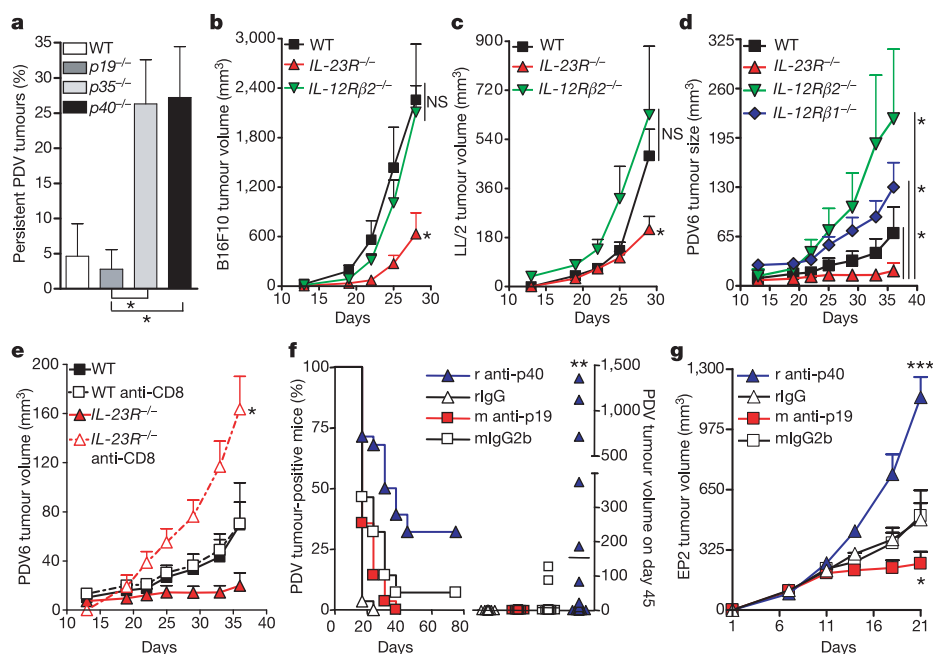


Figure 4 | Functional ablation of IL-23p19 signalling restricts tumour growth. **a**, Persistence of immune-sensitive PDV squamous carcinoma in IL-12-deficient but not WT or *p19*^{-/-} mice. **b**, B16F10 melanoma and **c**, LL/2 lung carcinoma cells exhibit reduced growth in *IL23R*^{-/-} mice. **d**, Tumour growth of immune-resistant PDV6 squamous carcinoma cells in IL-12- or IL-23-receptor-deficient mice. **e**, Growth restriction of PDV6 tumours is CD8-dependent only in *IL23R*^{-/-} mice. **f**, Antibody ablation

of IL-12p40 increases both tumour incidence and size compared to anti-IL-23p19 treatment. **g**, EP2 breast tumour growth is increased by ablation of IL-12p40 or reduced by anti-IL-23p19 treatment (****P* for days 18 and 21). Data are mean ± s.d.; scatter plots show incidence with bars depicting the mean. Significance of results in figures is represented by: **P* < 0.05 and ****P* < 0.001; NS, not significant; *n* ≥ 5 (**a–e**), *n* ≥ 14 (**f**) mice per group with two tumours per mouse for all studies.

nascent carcinomas (Fig. 4a). IL-23p19-deficient mice, however, had a strong resistance to tumour challenge, similar to that of C57BL/6 control mice. Although previous reports suggest genetically engineered tumours overexpressing an artificially linked IL-23p19 and p40 construct are rejected more efficiently²³, our results show that at the very least IL-23 deficiency does not impair the ability of the immune system to eliminate nascent tumours.

To explore the respective importance of the host's immune response to IL-12 and IL-23 and to control tumour growth, we challenged mice incapable of responding to IL-23 (*Il23R*^{-/-} mice) with subcutaneous transplantation of murine tumour cell lines. The growth of B16F10 melanoma and LL2 lung carcinoma was inhibited in *Il23R*^{-/-} mice incapable of responding to IL-23, compared to either WT or mice incapable of responding to IL-12 (Fig. 4b, c and data not shown). We then used an immune-resistant subclone of the PDV squamous carcinoma (termed PDV6), to investigate the role of IL-12 and IL-23 signalling in the host on tumour growth and CD8 T-cell function. Growth of squamous carcinomas in *Il23R*^{-/-} was reduced compared to that in WT mice, but increased in mice not able to respond to IL-12 (Fig. 4d). Although depletion of CD8 T cells in WT mice did not alter PDV6 tumour growth kinetics, it potentiated tumour growth in *Il23R*^{-/-} mice (Fig. 4e). These results demonstrate that the absence of IL-23 signalling attenuates tumour growth and suggests an essential role for CD8 T cells in restricting tumour growth in *Il23R*^{-/-} mice.

We next asked whether antibody ablation of IL-23 or combined ablation of IL-23 and IL-12 would have a different effect on the ability of the host immune system to eliminate early malignant tumour cells. Animals treated with anti-IL-23p19 showed a decreased risk of PDV tumour formation and faster elimination of injected tumour cells. However, animals treated with anti-IL-12p40, and thus depleted of both IL-12 and IL-23, had a significantly increased tumour incidence and developed larger, faster-growing tumours (Fig. 4f and data not shown). Finally, animals bearing tumours formed by a transformed mammary cell line were treated with IL-23 or IL-12 neutralizing antibodies. Mice treated with anti-IL-23p19 showed decreased tumour growth and increased tumour rejection. However, anti-IL-12p40 treatment led to a marked increase of tumour growth and an increase of metastasis formation (Fig. 4g and data not shown). Thus, although elimination of transformed cells is regulated by IL-12 (ref. 22), the contribution of the pro-inflammatory environment to tumour propagation can be seen in the result of functional IL-23 ablation.

The importance of inflammatory processes in tumour progression and initiation is exemplified by the tumour resistance of MMP9- or COX-2-deficient mice^{20,24}. However, local expression of growth factors and cytokines in tumours may not only aid tissue remodelling and angiogenesis, but also protect the tumour from immune-mediated elimination⁷. Surprisingly, our work shows that IL-23 appears not only to induce innate inflammatory infiltration, but also to downregulate dramatically the ability of CD8 T cells to infiltrate tumours (Supplementary Fig. S1). Although the understanding of the biological role of IL-23 is still evolving, it is known that IL-23 production by dendritic cells and macrophages can be elicited by bacterial exposure and Toll-like receptor engagement. The resulting IL-17-dependent immune response constitutes a significant component for rapid neutralization of infectious organisms^{9,25,26}. However, inappropriate expression of IL-23 also appears to coincide with numerous autoimmune disorders^{14,27}. IL-23-induced inflammatory processes such as induction of angiogenesis or macrophage and neutrophil infiltration are valuable defence mechanisms against bacterial infections. However, this regulatory pathway appears to be not just inappropriate for eliminating tumours but instead could provide a protective, tumour-promoting environment for nascent malignancies, suggesting that anti-IL-23p19 therapy may prove efficacious for tumour treatment.

METHODS

Animal models. C57BL/6 mice and *Il2p35*^{-/-} (*Il12a*^{tm1jm}), *Il2p40*^{-/-} (*Il12b*^{tm1jm}), *Il12Rβ1*^{-/-} (*Il12rb1*^{tm1jm}), and *Il12Rβ2*^{-/-} (*Il12rb2*^{tm1jm}) mice on the C57BL/6 background were obtained from Jackson Laboratory. *Il23R*^{-/-} (*Il23ra*^{tm1dnax}) were maintained on a C57BL/6 background; *Il23p19*^{-/-} (*Il23a*^{tm1dnax}) were produced by DNAX¹³ and made congenic on the C57BL/6 background (see Supplementary Methods). The two-step skin carcinogenesis model was adapted from ref. 28: mice were treated with 9,10-dimethyl-1,2-benzanthracene (DMBA; Sigma) in 200 μl acetone at 100 μg per mouse once at the age of 2–3 months, then treated twice-weekly with the tumour promoter 12-O-tetradecanoyl-phorbol acetate (TPA; Fisher) in 200 μl acetone, 30 μg per mouse for up to 1 year. Tumours observed arise as papillomas (keratoacanthomas) but statistically progressed towards carcinomas and metastasized via the lymph drainage. Papilloma counts were conducted routinely by visual examination. All experiments were repeated at least two times with similar significance values ($P = 0.0106$ and $P < 0.001$, respectively, $n \geq 7$ for each group in each experiment, with a total of $n \geq 21$ per genotype). Mice 8–10 weeks old were exposed to carcinogen for five months—they were either injected subcutaneously with murine IL-12, IL-23 or PBS at 20 ng per injection site or, for antibody ablation, systemically injected with 100 μg of anti-IL-23p19 antibody or isotype control. In both cases, skin was harvested 72 h later for immunohistochemistry or for TAQMAN gene expression analysis.

In tumour transplantation studies, transformed cell lines were mixed 1:1 with growth-factor-reduced Matrigel (Gibco) and injected subcutaneously or intradermally, typically in two separate sites per animal. Squamous carcinoma cells (PDV²⁹), immune resistant PDV derivative (PDV6; DNAX), melanoma (B16F10; ATCC) and lung (LL2; ATCC) cell lines were injected into C57BL/6 mice, while breast (EP2; DNAX and 4T1; ATCC) and colon (CT26; ATCC) cancer cells were injected into BALB/c mice. For persistence of PDV transplants, tumour growth was followed for an observation period of 2–6 months. Anti-cytokine antibody injections (1 mg per mouse) were repeated weekly, and expected serum concentrations were confirmed after one week and at the end of the experiment. For anti-CD8 antibody ablation, 0.25 μg per mouse of antibody (DNAX) was injected subcutaneously once per week; ablation effectiveness was judged by flow cytometry of treated spleens at the experimental endpoint. Tumour volume was determined by electronic calliper measure: (half-length)² × width.

Antibodies, cytokines, immunohistochemistry, zymogram. Recombinant murine IL-12 was obtained from R&D, while rmIL-23 was produced at DNAX as previously described⁹. For the ablation of IL-12, monoclonal antibody C17.8 (ref. 30) was used. Mouse anti-murine IL-23p19 antibodies were developed at DNAX. Antibodies were injected subcutaneously in equal PBS volume at 1 mg per mouse per week. Expected serum concentrations of C17.8 and anti-IL-23p19 were confirmed by immunoassays (ELISAs) specific for each antibody. Mouse skin biopsies were harvested from macroscopically tumour-free areas of carcinogen-exposed skin, snap-frozen, fixed and subjected to immunohistochemistry and counterstained with haematoxylin. Antibodies used include CD3, CD8, CD11b, CD11c, CD31 (BD Biosciences), active MMP9, perforin and F4/80 (R&D). Density of staining was determined by analysing multiple 20× fields with ImagePro Plus software (MediaCybernetics). Gelatin zymogram gels (Invitrogen) were performed following the supplier's instructions.

Gene expression analysis. For TAQMAN analysis, total RNA was isolated by standard methodologies and reverse transcribed. Expression analysis for marker-specific mRNA was measured using real-time quantitative PCR (ABI 5700) with SYBR Green PCR Mastermix (Applied Biosystems). Ubiquitin levels were measured in a separate reaction and used to normalize the data by the $\Delta - \Delta C_t$ method, using the mean cycle threshold (C_t) value for ubiquitin and the gene of interest for each sample; the expression $1.8^{(C_{t(\text{ubiquitin})} - C_{t(\text{gene of interest})})} \times 10^4$ was used to obtain the normalized values (see Supplementary Methods). For murine samples, punch biopsies were taken from at least three independent, macroscopically and histologically tumour-free areas for each condition.

Tumour-infiltrating immune cell isolation. Subcutaneous 4T1 and CT26 tumour masses were dissected and digested with 2 mg ml⁻¹ collagenase and 0.5 mg ml⁻¹ DNase (Sigma). Single-cell suspensions were created and live cells collected by centrifugation over a Histopaque (Sigma) gradient. Similar isolations from naive spleen were performed for comparison. Collected cells were positively selected for CD11b or CD11c by magnetic selection (Miltenyi Biotec) according to the manufacturer's specifications.

Statistical analyses. The following statistical analyses were used. For data in Fig. 1 and Supplementary Fig. S1, we used the Mann–Whitney test (non-paired Wilcoxon)—nonparametric, group comparison. In Figs 2, 3, 4a–c and Supplementary Figs S3 and S4, we used the Student's *t*-test. In Fig. 4a–c, e–g, we used the Kruskal–Wallis test followed by Dunn's multiple comparison test (non-parametric), or

analysis of variance (ANOVA) (parametric) followed by Tukey's multiple comparison test. In Fig. 4d, we used Wilcoxon matched pairs test. All tests (other than in Supplementary Fig. S2) were performed using Prism 4 (GraphPad Software). Linear regression analysis (Supplementary Fig. S2) was performed with Spotfire DecisionSite software.

Human tumour panels, immunohistochemistry. Paired human tumour and normal adjacent tissues were obtained from patients undergoing therapeutic or routine surgery, respectively (see Supplementary Methods). All sample diagnoses and staging were confirmed internally by a pathologist. Immunohistochemistry was performed on frozen sections with murine anti-human IL-23p19 (DNAX) with biotinylated horse anti-mouse IgG (Vector Laboratories).

Received 22 November 2005; accepted 11 April 2006.

Published online 10 May 2006.

1. Balkwill, F. & Mantovani, A. Inflammation and cancer: back to Virchow? *Lancet* **357**, 539–545 (2001).
2. Coussens, L. M. & Werb, Z. Inflammation and cancer. *Nature* **420**, 860–867 (2002).
3. Dunn, G. P., Old, L. J. & Schreiber, R. D. The immunobiology of cancer immunosurveillance and immunoediting. *Immunity* **21**, 137–148 (2004).
4. Yee, C. & Greenberg, P. Modulating T-cell immunity to tumours: new strategies for monitoring T-cell responses. *Nature Rev. Cancer* **2**, 409–419 (2002).
5. Blattman, J. N. & Greenberg, P. D. Cancer immunotherapy: a treatment for the masses. *Science* **305**, 200–205 (2004).
6. Marincola, F. M., Wang, E., Herlyn, M., Seliger, B. & Ferrone, S. Tumors as elusive targets of T-cell-based active immunotherapy. *Trends Immunol.* **24**, 335–342 (2003).
7. Ganss, R., Arnold, B. & Hammerling, G. J. Mini-review: overcoming tumor-intrinsic resistance to immune effector function. *Eur. J. Immunol.* **34**, 2635–2641 (2004).
8. Langrish, C. L. *et al.* IL-12 and IL-23: master regulators of innate and adaptive immunity. *Immunol. Rev.* **202**, 96–105 (2004).
9. Oppmann, B. *et al.* Novel p19 protein engages IL-12p40 to form a cytokine, IL-23, with biological activities similar as well as distinct from IL-12. *Immunity* **13**, 715–725 (2000).
10. Parham, C. *et al.* A receptor for the heterodimeric cytokine IL-23 is composed of IL-12R β 1 and a novel cytokine receptor subunit, IL-23R. *J. Immunol.* **168**, 5699–5708 (2002).
11. Harrington, L. E. *et al.* Interleukin 17-producing CD4⁺ effector T cells develop via a lineage distinct from the T helper type 1 and 2 lineages. *Nature Immunol.* **6**, 1123–1132 (2005).
12. Aggarwal, S., Ghilardi, N., Xie, M. H., de Sauvage, F. J. & Gurney, A. L. Interleukin-23 promotes a distinct CD4 T cell activation state characterized by the production of interleukin-17. *J. Biol. Chem.* **278**, 1910–1914 (2003).
13. Cua, D. J. *et al.* Interleukin-23 rather than interleukin-12 is the critical cytokine for autoimmune inflammation of the brain. *Nature* **421**, 744–748 (2003).
14. Langrish, C. L. *et al.* IL-23 drives a pathogenic T cell population that induces autoimmune inflammation. *J. Exp. Med.* **201**, 233–240 (2005).
15. Numasaki, M. *et al.* Interleukin-17 promotes angiogenesis and tumor growth. *Blood* **101**, 2620–2627 (2003).
16. Gupta, J. W., Kubin, M., Hartman, L., Cassatella, M. & Trinchieri, G. Induction of expression of genes encoding components of the respiratory burst oxidase during differentiation of human myeloid cell lines induced by tumor necrosis factor and gamma-interferon. *Cancer Res.* **52**, 2530–2537 (1992).
17. Girardi, M. *et al.* The distinct contributions of murine T cell receptor (TCR) $\gamma\delta^{+}$ and TCR $\alpha\beta^{+}$ T cells to different stages of chemically induced skin cancer. *J. Exp. Med.* **198**, 747–755 (2003).
18. Stark, M. A. *et al.* Phagocytosis of apoptotic neutrophils regulates granulopoiesis via IL-23 and IL-17. *Immunity* **22**, 285–294 (2005).
19. Mueller, M. M. *et al.* Autocrine growth regulation by granulocyte colony-stimulating factor and granulocyte macrophage colony-stimulating factor in human gliomas with tumor progression. *Am. J. Pathol.* **155**, 1557–1567 (1999).
20. Coussens, L. M., Tinkle, C. L., Hanahan, D. & Werb, Z. MMP-9 supplied by bone marrow-derived cells contributes to skin carcinogenesis. *Cell* **103**, 481–490 (2000).
21. Bergers, G. *et al.* Matrix metalloproteinase-9 triggers the angiogenic switch during carcinogenesis. *Nature Cell Biol.* **2**, 737–744 (2000).
22. Trinchieri, G. Interleukin-12 and the regulation of innate resistance and adaptive immunity. *Nature Rev. Immunol.* **3**, 133–146 (2003).
23. Lo, C. H. *et al.* Antitumor and antimetastatic activity of IL-23. *J. Immunol.* **171**, 600–607 (2003).
24. Tiano, H. F. *et al.* Deficiency of either cyclooxygenase (COX)-1 or COX-2 alters epidermal differentiation and reduces mouse skin tumorigenesis. *Cancer Res.* **62**, 3395–3401 (2002).
25. Lieberman, L. A. *et al.* IL-23 provides a limited mechanism of resistance to acute toxoplasmosis in the absence of IL-12. *J. Immunol.* **173**, 1887–1893 (2004).
26. Khader, S. A. *et al.* IL-23 compensates for the absence of IL-12p70 and is essential for the IL-17 response during tuberculosis but is dispensable for protection and antigen-specific IFN- γ responses if IL-12p70 is available. *J. Immunol.* **175**, 788–795 (2005).
27. Lee, E. *et al.* Increased expression of interleukin 23 p19 and p40 in lesional skin of patients with psoriasis vulgaris. *J. Exp. Med.* **199**, 125–130 (2004).
28. Nagase, H., Mao, J. H. & Balmain, A. A subset of skin tumor modifier loci determines survival time of tumor-bearing mice. *Proc. Natl Acad. Sci. USA* **96**, 15032–15037 (1999).
29. Fusenig, N. E., Am, S. M., Boukamp, P. & Worst, P. K. Characteristics of chemically transformed mouse epidermal cells *in vitro* and *in vivo*. *Bull. Cancer* **65**, 271–279 (1978).
30. Wysocka, M. *et al.* Interleukin-12 is required for interferon-gamma production and lethality in lipopolysaccharide-induced shock in mice. *Eur. J. Immunol.* **25**, 672–676 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Strader for her comments on the manuscript.

Author Contributions J.L.L. and X.Z. contributed equally to this work.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.O. (martin.oft@spcorp.com).

LETTERS

ATM stabilizes DNA double-strand-break complexes during V(D)J recombination

Andrea L. Bredemeyer¹, Girdhar G. Sharma², Ching-Yu Huang¹, Beth A. Helmink¹, Laura M. Walker¹, Katrina C. Khor¹, Beth Nuskey³, Kathleen E. Sullivan⁴, Tej K. Pandita², Craig H. Bassing^{3,5} & Barry P. Sleckman¹

The ATM (ataxia-telangiectasia mutated) protein kinase mediates early cellular responses to DNA double-strand breaks (DSBs) generated during metabolic processes or by DNA-damaging agents^{1–4}. ATM deficiency leads to ataxia-telangiectasia, a disease marked by lymphopenia, genomic instability and an increased predisposition to lymphoid malignancies with chromosomal translocations involving lymphocyte antigen receptor loci^{5,6}. ATM activates cell-cycle checkpoints and can induce apoptosis in response to DNA DSBs^{1–4}. However, defects in these pathways of the DNA damage response cannot fully account for the phenotypes of ATM deficiency. Here, we show that ATM also functions directly in the repair of chromosomal DNA DSBs by maintaining DNA ends in repair complexes generated during lymphocyte antigen receptor gene assembly. When coupled with the cell-cycle checkpoint and pro-apoptotic activities of ATM, these findings provide a molecular explanation for the increase in lymphoid tumours with translocations involving antigen receptor loci associated with ataxia-telangiectasia.

The exons encoding the variable region of lymphocyte antigen receptors are assembled through the process of V(D)J recombination⁷. This reaction occurs only in developing lymphocytes at the G1 phase of the cell cycle, and is initiated when the recombinase-activating gene-1 and -2 protein complex (hereafter referred to as RAG) introduces DNA DSBs at the border of two gene segments and their flanking recombination signals^{8–10}. DNA cleavage occurs only after the two recombination signals are brought together in a synaptic complex and leads to the formation of a pair of blunt phosphorylated signal ends and a pair of hairpin-sealed coding ends^{8–10}. The signal end and coding end pairs are processed and joined by non-homologous end-joining (NHEJ) DSB repair proteins to generate a signal join and coding join, respectively¹¹.

Although not absolutely required for V(D)J recombination, ATM can be recruited to RAG-induced DSBs, and the phenotypes of ATM-deficient humans and mice suggest that ATM may serve important functions during antigen receptor gene assembly^{5,6,12,13}. For example, thymic lymphomas with chromosomal translocations involving the T-cell receptor (TCR) α/δ locus in *Atm*^{-/-} mice are dependent on RAG expression^{14,15}. Similar defects are not observed in mice deficient in p53 or Chk2 proteins, which, upon activation by ATM, enforce the G1/S checkpoint and ultimately promote apoptosis if DNA DSBs are not repaired^{16,17}. Together, these findings suggest that ATM has additional undefined functions in the response to DNA DSBs generated during V(D)J recombination.

To elucidate potential ATM functions, we developed an experimental approach in which RAG-generated DNA DSBs can be induced, and their repair tracked, in G1-arrested wild-type and ATM-deficient cells. Viral (v)-Abl kinase-transformed pre-B-cell

lines were generated from multiple *Atm*^{+/+}, *Atm*^{+/-} and *Atm*^{-/-} mice expressing an E μ -Bcl2 transgene. Treatment of pre-B-cell lines with the Abl kinase inhibitor, STI571, leads to a rapid induction of RAG gene expression and a block in the G1-to-S transition that does not depend on RAG-induced DNA DSBs (Supplementary Fig. S1)¹⁸. Pre-B-cell lines were transduced with the pMX-RSS-GFP/IRES-hCD4 retroviral recombination substrate, hereafter referred to as pMX-INV (Fig. 1a)¹⁹. pMX-INV has a single pair of recombination signals that flank an anti-sense green fluorescent protein (GFP) cDNA and mediate recombination by inversion. Thus, normal rearrangement places the GFP cDNA in the sense orientation, allowing for GFP expression (Fig. 1a, b).

Pre-B-cell lines with pMX-INV integrants at diverse genomic sites (*Atm*^{+/+}:INV, *Atm*^{+/-}:INV and *Atm*^{-/-}:INV) were isolated by flow cytometric cell sorting for the retrovirally encoded hCD4 (Supplementary Fig. S2). Flow cytometric and Southern blotting analyses revealed that the level of normal pMX-INV rearrangement in STI571-treated *Atm*^{-/-}:INV cells was half that observed for *Atm*^{+/+}:INV or *Atm*^{+/-}:INV cells (Fig. 1b–d and Supplementary Fig. S3a, b). Analysis of pMX-INV rearrangements in *Atm*^{-/-}:INV and *Atm*^{+/+}:INV clones, and in independently derived *Atm*^{+/+} and *Atm*^{-/-} pre-B-cell lines, yielded comparable results (Supplementary Fig. S3a–c). Significant differences in cell viability or changes in the fraction of hCD4-expressing cells were not observed during the course of STI571 treatment (Supplementary Fig. S4). Together, these findings demonstrate that ATM deficiency leads to reduced levels of normal pMX-INV rearrangement that can not be attributed to defects in ATM-mediated G1/S checkpoint enforcement or apoptosis.

Southern blot analysis of STI571-treated *Atm*^{-/-}:INV pre-B cells, but not *Atm*^{+/+}:INV pre-B cells, revealed a novel band of the size expected for joining of the 3' coding end and 5' signal end of pMX-INV (HJ band, Fig. 1c). The joining of a signal end to the coding end generated by cleavage at the other recombination signal results in a hybrid join^{20,21}. Polymerase chain reaction (PCR) analyses confirmed an approximate 60-fold increase in hybrid join formation in STI571-treated *Atm*^{-/-}:INV cells, as compared to *Atm*^{+/+}:INV cells (Supplementary Fig. S3d, e). Decreased normal rearrangement and increased hybrid join formation were also observed in *Atm*^{+/-}:INV pre-B cells treated with STI571 and the ATM kinase-specific inhibitor KU-55933 (Supplementary Fig. S5)²². Hybrid join sequences were diverse, including nucleotide loss and the addition of both palindromic and non-templated nucleotides (Supplementary Fig. S6). Notably, the kinetics and level of hybrid join and coding join formation were similar in *Atm*^{-/-}:INV cells, demonstrating that, in the absence of ATM, hybrid join formation is a frequent outcome of pMX-INV rearrangement (Fig. 1c, d and Supplementary Fig. S3b). An increase in V κ 6-23 to J κ 1 hybrid joins was also observed in

¹Department of Pathology and Immunology and ²Department of Radiation Oncology, Washington University School of Medicine, St Louis, Missouri 63110, USA. ³Department of Pathology and Laboratory Medicine and ⁴Division of Allergy and Immunology, Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA. ⁵Abramson Family Cancer Research Institute, Philadelphia, Pennsylvania 19104, USA.

STI571-treated *Atm*^{-/-} pre-B-cell lines and in splenic B cells from *Atm*^{-/-} mice, as compared to their wild-type counterparts (Fig. 1e and Supplementary Fig. S7). Furthermore, increased hybrid joins involving V β 14 and V δ 5 were found in thymocytes from *Atm*^{-/-} mice (Supplementary Fig. S8). Thus, increased hybrid join formation

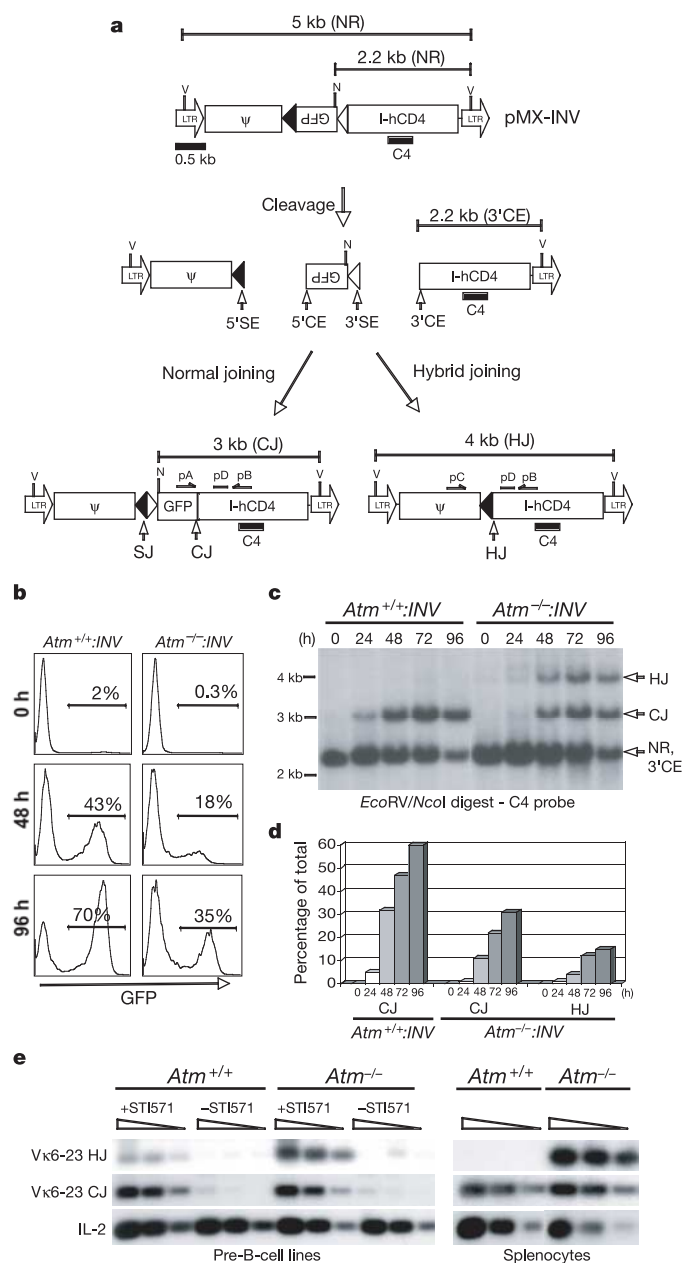


Figure 1 | ATM-deficient pre-B cells exhibit defects in V(D)J recombination. **a**, Schematic of non-rearranged (NR) pMX-INV, coding end (CE) and signal end (SE) DNA cleavage intermediates and signal join (SJ), coding join (CJ) and hybrid join (HJ) products. The long terminal repeats (LTR), packaging sequence (ψ), GFP cDNA, IRES-hCD4 cDNA (I-hCD4), 5' 12-recombination signal (12-RS) and 3' 23-RS (filled and open triangles, respectively), *EcoRV* (V) site, *NcoI* (N) site, the pA, pB, pC and pD oligonucleotides, and C4 probe are indicated. The recombination signals are not drawn to scale. **b–d**, *Atm*^{+/+}:INV and *Atm*^{-/-}:INV cells treated with STI571 for the indicated times were assayed for pMX-INV rearrangement by flow cytometry, with the percentage of GFP expressing cells indicated (**b**), or Southern blotting (**c**). The pMX-INV hybrid joins and coding joins in **c** are quantified in **d**. **e**, PCR analysis of V κ 6-23 to J κ 1 coding joins and hybrid joins in *Atm*^{+/+} and *Atm*^{-/-} pre-B-cell lines that were untreated (-STI571) or treated with STI571 for 96 h (+STI571), and in *Atm*^{+/+} and *Atm*^{-/-} splenocytes.

is a general feature of rearrangements that occur by inversion in *Atm*^{-/-} B and T cells.

Although some endogenous variable region exons are assembled by inversion, most proceed through rearrangements involving deletion of intervening sequences with chromosomal retention of coding joins. Thus, to determine whether hybrid join formation is a general feature of V(D)J recombination in *Atm*^{-/-} cells, we generated the pMX-DEL^{CJ} substrate (Fig. 2a). pMX-DEL^{CJ} differs from pMX-INV only by the inversion of the 5' recombination signal and its immediate flanking sequences and, therefore, undergoes rearrangement by deletion with chromosomal retention of the coding join (Fig. 2a). In contrast to pMX-INV rearrangements, pMX-DEL^{CJ} hybrid joins were not detected by Southern blot analyses of STI571-treated *Atm*^{-/-}:DEL^{CJ} pre-B cells (Fig. 2b). In agreement, PCR and flow cytometric analyses revealed comparably low levels of pMX-DEL^{CJ} hybrid joins in STI571-treated *Atm*^{+/+}:DEL^{CJ} and *Atm*^{-/-}:DEL^{CJ} pre-B cells (Fig. 2c and Supplementary Fig. S9). Hybrid joins involving endogenous V β gene segments that rearrange by deletion could not be detected by PCR in either *Atm*^{-/-} or *Atm*^{+/+} thymocytes (data not shown). Thus, ATM deficiency leads to an increase in hybrid join formation during rearrangements that occur by inversion, but not during those that occur by deletion.

How can the selective increase in hybrid join formation during rearrangement by inversion in ATM-deficient cells be explained? We considered the possibility that this may be due to a requirement for ATM in maintaining the stability of post-cleavage DNA end complexes. In this regard, loss of the intervening sequences (GFP cDNA) during pMX-INV rearrangement would leave the 5' signal end and 3' coding end that, if joined, would form a chromosomal hybrid join, whereas loss of this intervening sequence during pMX-DEL^{CJ} rearrangement would leave 5' and 3' coding ends that

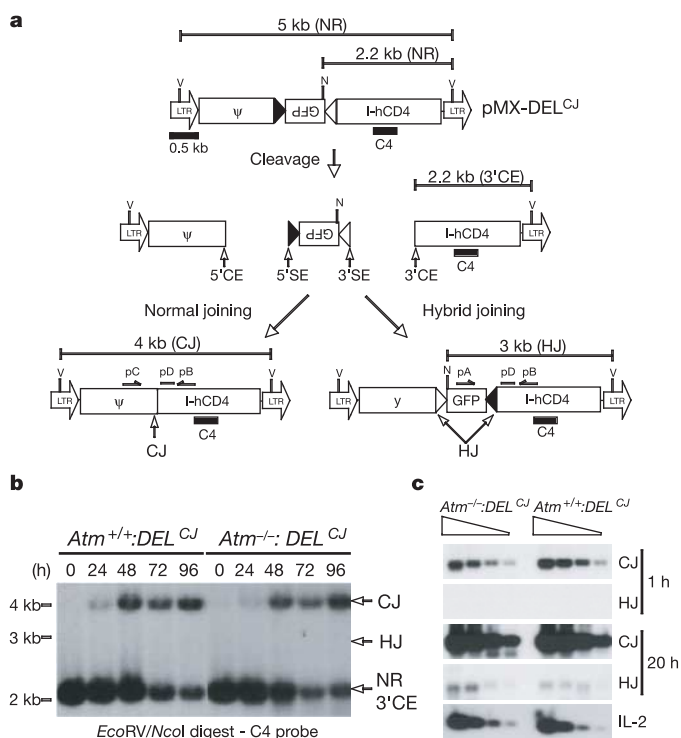


Figure 2 | Hybrid join formation is not increased during rearrangement by deletion. **a**, Schematic of pMX-DEL^{CJ} with components, intermediates and products as defined for pMX-INV in Fig. 1a. **b**, *Atm*^{+/+}:DEL^{CJ} and *Atm*^{-/-}:DEL^{CJ} cells treated with STI571 for the indicated times and pMX-DEL^{CJ} rearrangement assayed by Southern blotting. **c**, PCR analyses of pMX-DEL^{CJ} coding joins and hybrid joins in *Atm*^{+/+}:DEL^{CJ} and *Atm*^{-/-}:DEL^{CJ} cells treated for 96 h with STI571. Autoradiographs exposed for either 1 h or 20 h.

Several lines of evidence demonstrate that these bands represent unrepaired 3' coding ends. First, they are not present in cells after recovery from STI571 treatment, demonstrating that they are not due to stably joined products (R lanes, Fig. 3a, b). Second, a band of similar size is observed in STI571-treated pre-B cells derived from severe combined immunodeficient (SCID) mice, which have a defect in coding join formation and accumulate hairpin-sealed coding ends (Fig. 3a)²³. In contrast to SCID cells, this band becomes smaller and diffuse in size in *Atm*^{-/-} cells over time, suggestive of hairpin

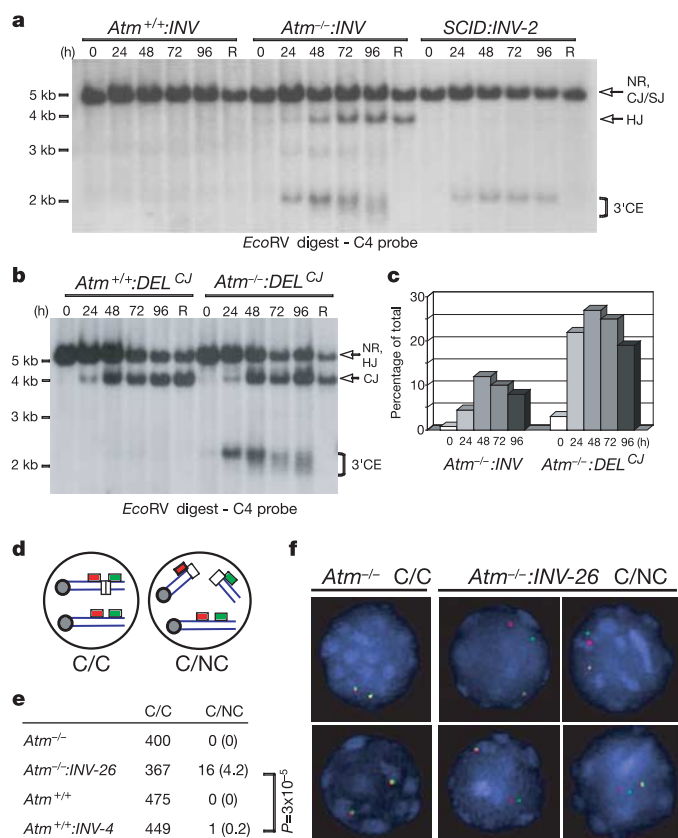


Figure 3 | Coding ends are lost from post-cleavage complexes in ATM-deficient cells. **a, b**, Southern blot analysis of genomic DNA from Fig. 1c (**a**) and Fig. 2b (**b**) and a SCID pre-B cell containing pMX-INV (*SCID:INV-2*). Lanes labelled R indicate cells treated with STI571 for 96 h followed by a 5-day recovery period. **c**, Quantification of 3' coding ends from **a** and **b**. **d**, Schematic of coincident (C) and non-coincident (NC) FISH probe hybridization on homologous chromosomes with (open rectangles) and without a pMX-INV integrant. The 5' and 3' probe sets are indicated as red and green rectangles, respectively. **e**, Number of C/C and C/NC nuclei after treatment with STI571 for 72 h. The percentage of C/NC nuclei are shown in parentheses. **f**, *Atm*^{-/-} nuclei showing C/C hybridization and *Atm*^{-/-}:INV-26 nuclei showing C/NC hybridization of the 5' probe set (labelled with Spectrum-Red) and the 3' probe set (labelled with Spectrum-Green).

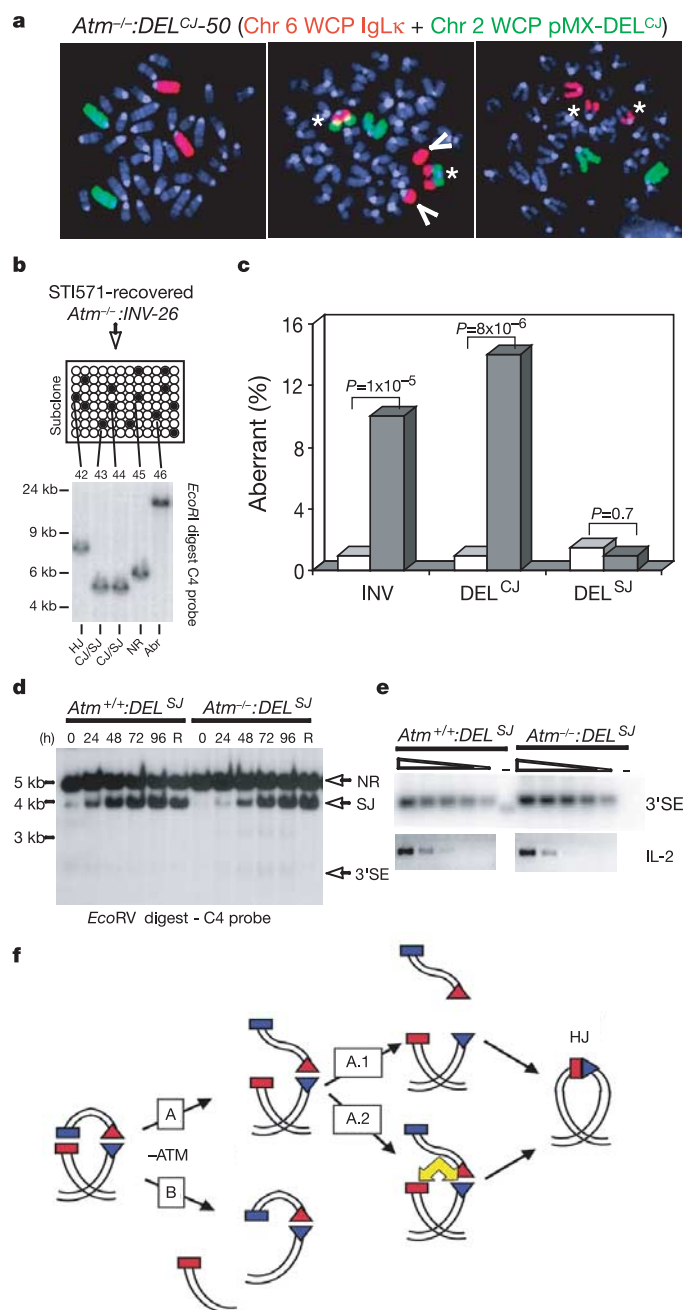


Figure 4 | Increased aberrant resolution of coding ends, but not signal ends, in ATM-deficient cells. **a**, Metaphase analysis of STI571-recovered *Atm*^{-/-}:*DEL*^{CJ}-50 cells using whole-chromosome paint to chromosome 6 (red), which contains the IgLk locus, and chromosome 2 (green), which contains pMX-DEL^{CJ}. Translocations (asterisks) and broken chromosomes (open arrows) are indicated. **b**, Experimental strategy for analysing pMX-INV rearrangements in clones of STI571-recovered *Atm*^{-/-}:*INV*-26 cells. Southern blot of clones with non-rearranged, normally rearranged (CJ/SJ), hybrid join and aberrant (Abr) pMX-INV rearrangements is shown. **c**, Fraction of aberrant pMX-INV (INV), pMX-DEL^{CJ} (DEL^{CJ}) and pMX-DEL^{SJ} (DEL^{SJ}) rearrangements in STI571-recovered *Atm*^{+/+} (white bars) and *Atm*^{-/-} (grey bars) clones. The primary data are in Supplementary Figs S16 and S19. **d**, **e**, Southern blot (**d**) and ligation-mediated PCR (**e**) analysis of pMX-DEL^{SJ} 3' signal ends in cells treated with STI571. **f**, Schematic of potential aberrant outcomes after RAG cleavage during rearrangements by inversion in ATM-deficient cells. The recombination signals (triangles) and coding segments (rectangles) paired before cleavage are the same colour. The intervening sequence referred to in the text is the segment of DNA flanked by the blue rectangle and red triangle.

opening and exonucleolytic processing (Fig. 3a, b). Third, a ligation-mediated PCR product of the expected size for a 3' coding end was observed only in STI571-treated *Atm*^{-/-}:INV and *Atm*^{-/-}:DEL^{CJ} cells, with the identity of this product verified by sequence analyses (Supplementary Fig. S10b, c and data not shown). Together, these findings demonstrate that the accumulation of unrepaired coding ends is a common outcome of V(D)J recombination in ATM-deficient cells, and that these coding ends are subject to hairpin opening and exonucleolytic processing.

To assess directly whether coding ends are lost from post-cleavage complexes in ATM-deficient cells, we carried out fluorescence *in situ* hybridization (FISH) on G1 nuclei from STI571-treated pre-B cells. To this end, pre-B-cell clones with single pMX-INV integrants were isolated (*Atm*^{-/-}:INV-26 and *Atm*^{+/-}:INV-4) (Supplementary Figs S5 and S10a). FISH probe sets spanning the pMX-INV integrants in *Atm*^{-/-}:INV-26 and *Atm*^{+/-}:INV-4 were generated and labelled with Spectrum-Red (5' probe set) or Spectrum-Green (3' probe set) (Supplementary Fig. S11a, b). Because the 5' and 3' probe sets are separated by short distances, the homologous chromosome (without pMX-INV) should yield overlapping, or coincident (C), hybridization (Fig. 3d). The chromosome with pMX-INV could yield non-coincident (NC) probe set hybridization if DNA ends have dissociated from post-cleavage complexes; otherwise this chromosome should yield coincident probe set hybridization. Therefore, we reasoned that two distinct hybridization patterns would be found, C/C or C/NC (Fig. 3d).

FISH analysis of nuclei from STI571-treated *Atm*^{-/-}:INV-26 cells revealed that 4.2% had a C/NC hybridization pattern (Fig. 3e, f). Southern blot analysis of these cells revealed that 15% had unrepaired pMX-INV 3' coding ends (data not shown). The difference in the fraction of nuclei with C/NC hybridization and those with 3' coding ends could be due, in part, to DNA ends that have dissociated at distances not resolved by FISH. The parental *Atm*^{-/-} pre-B-cell line, which does not contain a pMX-INV integrant, yielded only the C/C probe hybridization pattern, and the *Atm*^{+/-}:INV-4 clone yielded only 1 nucleus (of 450) with a C/NC hybridization pattern (Fig. 3e, f and Supplementary Fig. S11c). Similar results were observed upon analyses of G1 nuclei from *Atm*^{-/-}:DEL^{CJ} and *Atm*^{+/-}:DEL^{CJ} clones (Supplementary Figs S12 and S13).

If the NC hybridization pattern is due to dissociated broken DNA ends, passage of these cells into the S phase should lead to the generation of replicated chromosomal breaks, as was observed for DSBs generated during immunoglobulin class switch recombination in proliferating ATM-deficient B cells²⁴. Indeed, metaphase FISH analysis of *Atm*^{-/-}:INV and *Atm*^{-/-}:DEL^{CJ} clones released from STI571 treatment (hereafter referred to as STI571-recovered cells) revealed that approximately 2% contained replicated chromosomal breaks spanning the retroviral integrant (Supplementary Fig. S14). Such replicated chromosomal breaks were not observed in STI571-recovered parental *Atm*^{-/-} cells or in *Atm*^{+/-}:INV and *Atm*^{+/-}:DEL^{CJ} clones (Supplementary Fig. S14). Together, these findings demonstrate that coding ends can become dissociated from post-cleavage complexes generated during V(D)J recombination in G1-phase ATM-deficient cells.

One of the hallmarks of ataxia-telangiectasia is the generation of aberrant rearrangements during V(D)J recombination, such as translocations, large chromosomal deletions and inversions^{5,6}. We took several approaches to assay for similar aberrant rearrangements involving retroviral recombination substrates in STI571-recovered *Atm*^{-/-} pre-B cells. Metaphase FISH analyses of STI571-recovered *Atm*^{-/-}:DEL^{CJ} clones revealed that approximately 3% had karyotypic abnormalities, including replicated chromosomal breaks, translocations and complex rearrangements occurring at the pMX-DEL^{CJ} integration site (Fig. 4a and Supplementary Figs S14 and S15). Furthermore, whole-chromosome paint analyses revealed reciprocal translocations between chromosome 6 (which contains the IgLk locus) and chromosomes harbouring pMX-DEL^{CJ}, as well as other

chromosomes (Fig. 4a and Supplementary Fig. S15). These abnormalities were rarely observed in untreated *Atm*^{-/-}:DEL^{CJ} or STI571-recovered *Atm*^{+/-}:DEL^{CJ} clones (Supplementary Figs S14 and S15).

The full spectrum of aberrant rearrangements in ATM-deficient cells may not be detected through cytogenetic analyses such as those described above. Accordingly, pMX-INV and pMX-DEL^{CJ} rearrangements were assayed by Southern blotting of clones of STI571-recovered cells (Fig. 4b, Supplementary Fig. S16a and data not shown). The fraction of clones with normal or hybrid join rearrangements was in agreement with pMX-INV and pMX-DEL^{CJ} rearrangement in bulk pre-B-cell populations (Supplementary Fig. S16b, c). Notably, 10–14% of the STI571-recovered *Atm*^{-/-}:INV-26 and *Atm*^{-/-}:DEL^{CJ}-50 clones had aberrant rearrangements, whereas only 1% of the STI571-recovered *Atm*^{+/-}:INV-4 and *Atm*^{+/-}:DEL^{CJ}-119 clones had such events (Fig. 4c and Supplementary Fig. S16b, c). Characterization of aberrant pMX-INV and pMX-DEL^{CJ} rearrangements revealed that they were heterogeneous in nature and included large deletions and inversions involving the retroviral substrate, as well as translocations between the retroviral substrate and the IgLk locus (Supplementary Figs S17 and S18).

To address whether ATM is required to maintain signal end complexes, we generated the pMX-DEL^{SJ} retroviral recombination substrate (Supplementary Fig. S19a). pMX-DEL^{SJ} is identical to pMX-DEL^{CJ} except that the recombination signals have been inverted such that rearrangement, which occurs by deletion, leaves the signal join in the chromosome (Supplementary Fig. S19a). Like pMX-DEL^{CJ}, pMX-DEL^{SJ} underwent robust rearrangement in *Atm*^{-/-}:DEL^{SJ} cells, forming a high frequency of precise signal joins (Fig. 4d and Supplementary Fig. 19b). In contrast to coding ends in *Atm*^{-/-}:DEL^{CJ} cells, signal end accumulation was not observed during pMX-DEL^{SJ} rearrangement in *Atm*^{-/-}:DEL^{SJ} pre-B cells (Fig. 4d, e). Significant signal end accumulation was observed during pMX-DEL^{SJ} rearrangement in *Ku70*^{-/-} pre-B cells (data not shown).

It is possible that signal ends do not accumulate in ATM-deficient cells because they are rapidly resolved as aberrant rearrangements after their loss from post-cleavage complexes. However, analysis of STI571-recovered *Atm*^{-/-}:DEL^{SJ} clones revealed that aberrant pMX-DEL^{SJ} rearrangements are rare and occur at a frequency similar to that observed in STI571-recovered *Atm*^{+/-}:DEL^{SJ} clones (Fig. 4c and Supplementary Fig. S19c). These findings demonstrate that, unlike coding ends, ATM function is not required to maintain signal ends in post-cleavage complexes. In this regard, RAG has been shown to bind avidly to signal end pairs, forming stable signal end complexes after cleavage *in vitro*, and thus may perform the same function during chromosomal V(D)J recombination *in vivo*^{25,26}.

Here we demonstrate that ATM functions in the repair of chromosomal DNA DSBs generated during V(D)J recombination. The accumulation of unrepaired coding ends in ATM-deficient cells could reflect a requirement for ATM in the recruitment and/or function of NHEJ proteins, as has been suggested for the repair of some DSBs generated by DNA-damaging agents²⁷. However, the increase in hybrid join formation during rearrangement by inversion cannot be solely attributed to a defect in NHEJ, because these activities are normally required for efficient hybrid join formation in cells expressing full-length RAG proteins²⁸. It is possible that the accumulation of coding ends and increase in hybrid joins result from defects in distinct ATM-dependent pathways. In this regard, truncated RAG proteins can catalyse hybrid join formation by transposition; thus, ATM could modulate the transposition capability of full-length RAG proteins *in vivo*^{8–10,28}. However, this activity would have to be specific for rearrangements by inversion, as increased hybrid joins were not observed in ATM-deficient cells during rearrangements by deletion.

Alternatively, the coding end accumulation and increase in hybrid joins could reflect a function for ATM in maintaining coding ends in post-cleavage complexes, either by preventing coding end loss or by rapidly reclaiming released coding ends (Fig. 4f). Dissociated coding

ends would accumulate in ATM-deficient cells and may participate in aberrant rearrangements such as translocations (Fig. 4f, pathway B). Coupling normal signal end dissociation with increased coding end dissociation, during rearrangement by inversion, could lead to loss of the intervening sequence, with the remaining two chromosomal ends (coding end and signal end) forming a hybrid join (Fig. 4f pathway A.1). In contrast, during rearrangement by deletion, loss of the intervening sequence leaves two chromosomal coding ends, which would form a coding join (Fig. 2a). It is also possible that an unpaired coding end may be able to disrupt a signal end complex exclusively during rearrangements by inversion, leading to the formation of a hybrid join (Fig. 4f, pathway A.2). ATM could stabilize coding ends in post-cleavage complexes directly or through the activation of downstream targets.

Our findings provide a potential molecular explanation for the increased incidence of lymphoid tumours with antigen receptor locus translocations in ataxia-telangiectasia. Compound deficiencies in NHEJ and p53 proteins lead to a marked increase in the incidence of lymphoid tumours with antigen receptor locus translocations, whereas isolated deficiencies in either NHEJ proteins or p53 do not²⁹. Thus, the generation of these oncogenic lesions requires both defects in DSB repair (NHEJ deficiency) and defects in DSB-induced checkpoints/apoptosis (p53 deficiency)²⁹. Here we show that, in addition to checkpoint activation, ATM is also directly involved in the repair of RAG-induced DSBs. Thus, the singular deficiency of ATM would result in compound defects in both DSB-induced checkpoints/apoptosis and DSB repair, resulting in a predisposition to lymphomas with chromosomal translocations involving antigen receptor loci. By extension, we hypothesize that this DSB repair function of ATM also leads to the general genomic instability phenotype of ataxia telangiectasia and contributes to the suppression of oncogenic translocations from different types of DNA DSBs induced in a wide variety of cells.

METHODS

Generation of retroviral recombination substrates. The pMX-INV, pMX-DEL^{CJ} and pMX-DEL^{SJ} retroviral recombination substrates were generated as described in Supplementary Methods. These retroviral substrates all contain GFP cDNA (GFP) and human CD4 cDNA downstream of an internal ribosomal entry site (IRES). They differ from each other only with respect to the orientation of the two recombination signals.

Cell culture. v-Abl-transformed pre-B cell lines, derived from mice that contained the E μ -Bcl2 transgene and different gene mutations, were generated as described in Supplementary Methods. The A70 and 2A pre-B-cell lines were generated from different *Atm*^{+/+} mice. The 2E and 2F pre-B-cell lines were generated from the same *Atm*^{-/-} mouse whereas the A72 pre-B-cell line was generated from a separate *Atm*^{-/-} mouse.

Southern and northern blot analyses and PCR. These analyses were carried out as described in Supplementary Methods. PCR analysis of hybrid join, coding join and signal join formation in the retroviral recombination substrates was carried out using a single amplification. V β 14 and V δ 5 coding join formation was assayed using a single amplification, whereas V β 14 and V δ 5 hybrid joins were assayed by sequential nested PCR amplification reactions. V κ 6-23 hybrid joins and coding joins were assayed by sequential nested PCR amplification reactions. PCR of the IL-2 gene is provided in all cases to control for the quantity of template DNA. All PCR titrations are serial fourfold dilutions.

Fluorescence in situ hybridization. BACs or PCR fragments were generated for use as probes, as described in Supplementary Methods. Probes were labelled with Spectrum-Red or Spectrum-Green. Mouse whole-chromosome paints were obtained from Applied Spectral Imaging Inc. Analysis was carried out using a Zeiss Axioplan2ie epifluorescence microscope with ISIS image acquisition software.

Received 5 January; accepted 3 May 2006.

Published online 14 June 2006.

1. Kastan, M. B. & Lim, D. S. The many substrates and functions of ATM. *Nature Rev. Mol. Cell Biol.* **1**, 179–186 (2000).
2. Pandita, T. K. A multifaceted role for ATM in genome maintenance. *Expert Rev. Mol. Med.* **5**, 1–21 (2003).
3. Shiloh, Y. ATM and related protein kinases: safeguarding genome integrity. *Nature Rev. Cancer* **3**, 155–168 (2003).

4. Durocher, D. & Jackson, S. P. DNA-PK, ATM and ATR as sensors of DNA damage: variations on a theme? *Curr. Opin. Cell Biol.* **13**, 225–231 (2001).
5. Lavin, M. F. & Shiloh, Y. The genetic defect in ataxia-telangiectasia. *Annu. Rev. Immunol.* **15**, 177–202 (1997).
6. Taylor, A. M., Metcalfe, J. A., Thick, J. & Mak, Y. F. Leukemia and lymphoma in ataxia telangiectasia. *Blood* **87**, 423–438 (1996).
7. Tonegawa, S. Somatic generation of antibody diversity. *Nature* **302**, 575–581 (1983).
8. Fugmann, S. D., Lee, A. I., Shockett, P. E., Willey, I. J. & Schatz, D. G. The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu. Rev. Immunol.* **18**, 495–527 (2000).
9. Oettinger, M. A. V(D)J recombination: on the cutting edge. *Curr. Opin. Cell Biol.* **11**, 325–329 (1999).
10. Gellert, M. V(D)J recombination: rag proteins, repair factors, and regulation. *Annu. Rev. Biochem.* **71**, 101–132 (2002).
11. Bassing, C. H. & Alt, F. W. The cellular response to general and programmed DNA double strand breaks. *DNA Repair (Amst.)* **3**, 781–796 (2004).
12. Matei, I. R., Guidos, C. J. & Danska, J. S. ATM-dependent DNA damage surveillance in T-cell development and leukemogenesis: the DSB connection. *Immunol. Rev.* **209**, 142–158 (2006).
13. Perkins, E. J. et al. Sensing of intermediates in V(D)J recombination by ATM. *Genes Dev.* **16**, 159–164 (2002).
14. Liao, M. J. & Van Dyke, T. Critical role for Atm in suppressing V(D)J recombination-driven thymic lymphoma. *Genes Dev.* **13**, 1246–1250 (1999).
15. Petiniot, L. K. et al. Recombinase-activating gene (RAG) 2-mediated V(D)J recombination is not essential for tumorigenesis in Atm-deficient mice. *Proc. Natl Acad. Sci. USA* **97**, 6664–6669 (2000).
16. Donehower, L. A. et al. Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* **356**, 215–221 (1992).
17. Hiraio, A. et al. Chk2 is a tumor suppressor that regulates apoptosis in both an ataxia telangiectasia mutated (ATM)-dependent and an ATM-independent manner. *Mol. Cell Biol.* **22**, 6521–6532 (2002).
18. Muljo, S. A. & Schlissel, M. S. A small molecule Abl kinase inhibitor induces differentiation of Abelson virus-transformed pre-B cell lines. *Nature Immunol.* **4**, 31–37 (2003).
19. Liang, H. E. et al. The “dispensable” portion of RAG2 is necessary for efficient V-to-DJ rearrangement during B and T cell development. *Immunity* **17**, 639–651 (2002).
20. Lewis, S. M., Hesse, J. E., Mizuuchi, K. & Gellert, M. Novel strand exchanges in V(D)J recombination. *Cell* **55**, 1099–1107 (1988).
21. Morzycka-Wroblewska, E., Lee, F. E. & Desiderio, S. V. Unusual immunoglobulin gene rearrangement leads to replacement of recombinational signal sequences. *Science* **242**, 261–263 (1988).
22. Hickson, I. et al. Identification and characterization of a novel and specific inhibitor of the ataxia-telangiectasia mutated kinase ATM. *Cancer Res.* **64**, 9152–9159 (2004).
23. Roth, D. B., Menetski, J. P., Nakajima, P. B., Bosma, M. J. & Gellert, M. V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in scid mouse thymocytes. *Cell* **70**, 983–991 (1992).
24. Franco, S. et al. H2AX prevents DNA breaks from progressing to chromosome breaks and translocations. *Mol. Cell* **21**, 201–214 (2006).
25. Agrawal, A. & Schatz, D. G. RAG1 and RAG2 form a stable postcleavage synaptic complex with DNA containing signal ends in V(D)J recombination. *Cell* **89**, 43–53 (1997).
26. Hiom, K. & Gellert, M. Assembly of a 12/23 paired signal complex: a critical control point in V(D)J recombination. *Mol. Cell* **1**, 1011–1019 (1998).
27. Riballo, E. et al. A pathway of double-strand break rejoining dependent upon ATM, Artemis, and proteins locating to γ -H2AX foci. *Mol. Cell* **16**, 715–724 (2004).
28. Sekiguchi, J. A., Whitlow, S. & Alt, F. W. Increased accumulation of hybrid V(D)J joins in cells expressing truncated versus full-length RAGs. *Mol. Cell* **8**, 1383–1390 (2001).
29. Mills, K. D., Ferguson, D. O. & Alt, F. W. The role of DNA breaks in genomic instability and tumorigenesis. *Immunol. Rev.* **194**, 77–95 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Schlissel for pMX-RSS-GFP/IRES-hCD4. STI571 was provided by Novartis Pharmaceuticals and KU-55933 was provided by G. Smith at KuDOS Pharmaceuticals. We thank H. Virgin, K. Murphy and M. Krangel for critical review of the manuscript. This work is supported in part by grants from the National Institutes of Health (to B.P.S. and T.K.P.) and the Department of Army (T.K.P.). C.H.B. is a Pew Scholar in the Biomedical Sciences and a recipient of a Pilot Project Award from the American Cancer Society. B.P.S. is a recipient of a Research Scholar Award from the American Cancer Society. C.-Y.H. is supported by a post-doctoral training grant and A.L.B. a pre-doctoral training grant from the NIH.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.P.S. (Sleckman@immunology.wustl.edu).

Degradation of Id2 by the anaphase-promoting complex couples cell cycle exit and axonal growth

Anna Lasorella^{1,2,3}, Judith Stegmüller⁵, Daniele Guardavaccaro⁶, Guangchao Liu¹, Maria S. Carro¹, Gerson Rothschild¹, Luis de la Torre-Ubieta⁵, Michele Pagano⁶, Azad Bonni⁵ & Antonio Iavarone^{1,2,4}

In the developing nervous system, Id2 (inhibitor of DNA binding 2, also known as inhibitor of differentiation 2) enhances cell proliferation, promotes tumour progression and inhibits the activity of neurogenic basic helix–loop–helix (bHLH) transcription factors^{1,2}. The anaphase promoting complex/cyclosome and its activator Cdh1 (APC/C^{Cdh1}) restrains axonal growth but the targets of APC/C^{Cdh1} in neurons are unknown^{3–5}. Id2 and other members of the Id family are very unstable proteins that are eliminated as cells enter the quiescent state, but how they are targeted for degradation has remained elusive^{6,7}. Here we show that Id2 interacts with the core subunits of APC/C and Cdh1 in primary neurons. APC/C^{Cdh1} targets Id2 for degradation through a destruction box motif (D box) that is conserved in Id1 and Id4. Depletion of Cdh1 stabilizes Id proteins in neurons, whereas Id2 D-box mutants are impaired for Cdh1 binding and remain stable in cells that exit from the cell cycle and contain active APC/C^{Cdh1}. Mutants of the Id2 D box enhance axonal growth in cerebellar granule neurons *in vitro* and in the context of the cerebellar cortex, and overcome the myelin inhibitory signals for growth. Conversely, activation of bHLH transcription factors induces a cluster of genes with potent axonal inhibitory functions including the gene coding for the Nogo receptor, a key transducer of myelin inhibition. Degradation of Id2 in neurons permits the accumulation of the Nogo receptor, thereby linking APC/C^{Cdh1} activity with bHLH target genes for the inhibition of axonal growth. These findings indicate that deregulated Id activity might be useful to reprogramme quiescent neurons into the axonal growth mode.

To identify protein complexes of Id2 in human neuroblastoma cells, we used immunoaffinity chromatography followed by tandem mass spectrometry. New Id2 partners are the APC/C subunits Apc1, Apc5 and Apc8/Cdc23 (Supplementary Fig. 1a). The identification *in vivo* of complexes of Id2 and subunits of APC/C indicates that Id2 might be targeted for degradation by APC/C. The enzymatic E3 ubiquitin ligase activity of APC/C requires either the Cdc20 or Cdh1 co-activator⁴. Expression of Cdh1 but not that of Cdc20 led to a marked decrease in Id2 level; this effect was prevented by proteasomal inhibition (Fig. 1a). Moreover, Id2 was eliminated at a faster rate in the presence of Cdh1 (Fig. 1b). Expression of Cdh1 decreased endogenous Id2 in asynchronously growing U2OS cells without changing the cell cycle distribution, and also in cells synchronized in the S phase of the cell cycle with aphidicolin (Fig. 1c, d, and Supplementary Fig. 1b, c). Activation of APC/C^{Cdh1} also eliminated E47-bound Id2 and destabilized Id2 carrying a deletion of the amino-terminal 15 amino-acid residues, which have been shown to contribute to Id2 ubiquitination⁷ (Supplementary Fig. 1d, e). Id2 associated with Cdh1 in the absence of proteasomal inhibition, but the interaction was more efficient in cells treated with MG-132

(Supplementary Fig. 1f, g). To examine whether Id2 becomes unstable in cells undergoing quiescence we monitored Id2 expression in response to serum deprivation. Id2 was downregulated and underwent accelerated degradation in NIH 3T3 fibroblasts deprived of serum mitogens (Supplementary Fig. 2a, b). We observed similar results in serum-starved U2OS-Id2 and SK-N-SH neuroblastoma cells undergoing arrest in G0/G1 after treatment with retinoic acid (Supplementary Fig. 3a–c)⁸. To examine whether APC/C^{Cdh1} regulates Id2 stability *in vivo* we studied the levels of Id2 after a decrease in APC/C activity through the depletion of Cdh1. Knockdown of Cdh1 by RNA interference led to elevation of the steady-state levels of

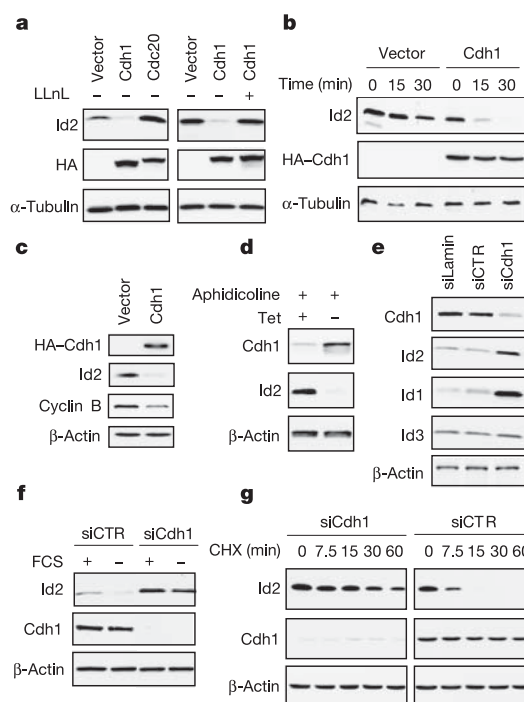


Figure 1 | Id2 is a substrate of APC/C^{Cdh1}. **a**, HeLa cells were transfected with Id2, haemagglutinin (HA)-conjugated Cdh1 or HA-Cdc20 and treated with N-Acetyl-L-leucyl-L-leucyl-L-norleucinal (LLNL). **b**, Transfected HeLa cells were treated with cycloheximide. **c**, Endogenous Id2 in U2OS cells transfected with HA-Cdh1. **d**, U2OS-tet-Cdh1 cells were treated with aphidicolin before the removal of tetracycline. **e**, **f**, LAN-1 neuroblastoma cells (**e**) and U2OS-Id2 cells (**f**) transfected with siRNA oligonucleotides. **g**, LAN-1 cells were transfected with siRNA; quiescence was induced by retinoic acid and cells were treated with cycloheximide (CHX).

¹Institute for Cancer Genetics, and Departments of ²Pathology, ³Pediatrics and ⁴Neurology, College of Physicians and Surgeons of Columbia University, New York, New York 10032, USA. ⁵Department of Pathology, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. ⁶Department of Pathology, NYU Cancer Institute, New York University School of Medicine, 550 First Avenue MSB 599, New York, New York 10016, USA.

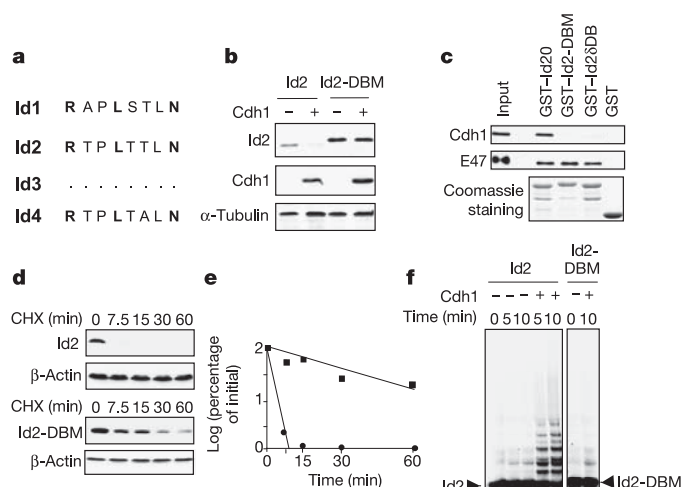


Figure 2 | D-box-dependent degradation of Id proteins by APC/C^{Cdh1}. **a**, Alignment of D boxes in Id proteins. **b**, HeLa cells were transfected with Id2 or Id2-DBM and HA-Cdh1 and analysed by western blotting. **c**, Pull-down assay with GST fusion proteins and Cdh1 and E47 translated *in vitro*. **d**, HeLa cells were transfected with Id2 or Id2-DBM and treated with cycloheximide (CHX). **e**, Quantification of Id2 (circles) and Id2-DBM (squares) from **d**. **f**, *In vitro* ubiquitination assay of Id2 and Id2-DBM by immunopurified APC/C in the presence or absence of Cdh1 translated *in vitro*. Arrowheads, unmodified Id2/Id2-DBM.

endogenous Id2 and Id1 in LAN1 neuroblastoma cells and exogenous Id2 in U2OS cells (Fig. 1e, and Supplementary Fig. 3d). It also prevented the downregulation of Id2 in quiescent U2OS cells and led to a tenfold increase in the Id2 half-life in LAN1 cells undergoing cell cycle arrest after treatment with retinoic acid (Fig. 1f, g, and Supplementary Fig. 3e, f). However, Id3 was unchanged in Cdh1-depleted cells (Fig. 1e). In support of a role of APC/C^{Cdh1} in Id2 stability, expression of Emi1, a specific inhibitor of APC/C (ref. 9), led to the accumulation of Id2 (Supplementary Fig. 3g).

We noted that Id2 contains a canonical D-box motif (RxxLxxxN)¹⁰ at residues 100–107, indicating that Id2 might be a direct substrate of APC/C^{Cdh1}. The D-box motif is conserved in Id1 and Id4 but not in Id3 (Fig. 2a). We generated mutant Id2 proteins in which the key arginine and leucine residues were changed into glycine and valine, respectively (Id2-DBM) and the D box was deleted (Id2δDB). Id2-DBM and Id2δDB mutants were expressed at levels higher than wild-type Id2 and were resistant to Cdh1-mediated destabilization (Fig. 2b, and Supplementary Fig. 4a). Conversely, APC/C^{Cdh1} eliminated the Id2 mutant that lacks the HLH region (Id2δHLH) and also wild-type Id1 and Id4 but not Id3 (Supplementary Fig. 4b–e). Next we investigated whether the D box is essential for the recognition of Id2 by APC/C^{Cdh1}. Glutathione S-transferase (GST)–Id2 but not GST–Id2-DBM or GST–Id2δDB captured *in vitro* translated Cdh1 and Cdh1 from HeLa cell extract, whereas all GST–Id2 fusion proteins captured Apc1 and Cdc27 (Fig. 2c, and Supplementary Fig. 5a). Similarly, both Flag–Id2 and Flag–Id2-DBM associated with APC/C core subunits, but Flag–Id2-DBM precipitated much less

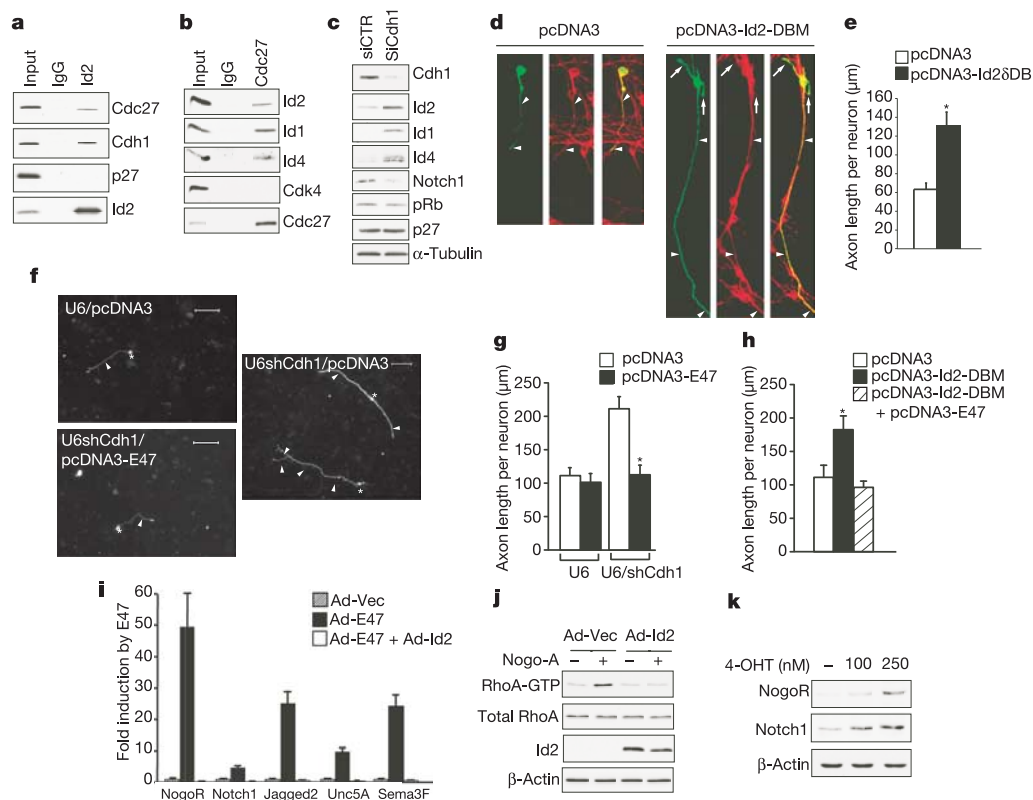


Figure 3 | An APC/C^{Cdh1}–Id–BHLH pathway controls axonal growth. **a**, **b**, Immunoprecipitation of Id2 (**a**) and Cdc27 (**b**) from CGNs. **c**, CGNs transfected with siRNA oligonucleotides and analysed by western blotting. Notch1 is the Val 1744-cleaved intracellular fragment. **d**, GFP (green) and Tau (red) staining in CGNs transfected with vector and Id2-DBM. **e**, Axonal length measured in CGNs transfected with vector or Id2δDB together with GFP ($n = 119$; asterisk, $P \leq 0.0001$). Results are means $\pm$ s.e.m. **f**, CGNs transfected with E47 and U6-CTR or U6-shCdh1 together with GFP. **g**, Quantification of axonal length from **f** ($n = 300$; asterisk, $P \leq 0.001$).

Results are means $\pm$ s.e.m. **h**, Axonal length measured from CGNs transfected with Id2-DBM in the absence or presence of E47 ($n = 76$; asterisk, $P \leq 0.01$). Results are means $\pm$ s.e.m. **i**, Expression of E47 target genes in SK-N-SH neuroblastoma cells infected with adeno-vector, adeno-E47 and adeno-E47 plus adeno-Id2. NogoR, Nogo receptor. Results are means $\pm$ s.d. ($n = 3$). **j**, RhoA activation assay in adenovirus-infected SK-N-SH cells treated with Nogo-A peptide. **k**, SF210-E47-ER cells treated with the indicated concentrations of 4-OHT.

Cdh1 than Flag-Id2 (Supplementary Fig. 5b). To determine the importance of the D-box-mediated degradation of Id2 we monitored the rate of degradation of Id2 and Id2-DBM. In transiently transfected HeLa cells the half-life of Id2-DBM was extended more than tenfold compared with wild-type Id2 (Fig. 2d, e). In addition, treatment of U2OS cells with the antimitogenic cytokine transforming growth factor- β efficiently depleted Id2 but did not decrease Id2-DBM (Supplementary Fig. 5c). Accordingly, wild-type Id2 but not Id2-DBM was polyubiquitinated *in vitro* by APC/C in the presence, but not in the absence, of Cdh1 (Fig. 2f, and Supplementary Fig. 5d). Together, these data indicate that APC/C^{Cdh1} can polyubiquitinate Id2 in a D-box-dependent manner. In contrast, Id2 did not affect the integrity of the APC/C complex or impair the ability of APC/C^{Cdh1} to degrade other substrates (Supplementary Fig. 6).

Apart from the regulation of the cell cycle, the expression of APC/C^{Cdh1} in postmitotic neurons has been puzzling¹¹. Previous studies indicated that APC/C^{Cdh1} directs cell-cycle-independent functions in postmitotic neurons, including the negative control of axonal growth. At embryonic day 16 (E16), Id2 was physically associated with brain APC/C (Supplementary Fig. 7a). We therefore examined whether a complex containing Id2 and APC/C^{Cdh1} exists in differentiated neurons, and whether APC/C^{Cdh1} directs the degradation of Id2 to control axonal growth. We found that lysates from primary cerebellar granule neurons (CGNs) and differentiated neuroblastoma cells contained Id2-core APC/C and Id2-Cdh1 complexes (Fig. 3a, b, and Supplementary Fig. 7d, e). Furthermore, postmitotic neurons transfected with Cdh1 short interfering RNA (siRNA) accumulated Id2, Id1 and Id4 in the absence of cell cycle re-entry as indicated by the lack of bromodeoxyuridine incorporation, retinoblastoma gene product phosphorylation or decrease in the cyclin-dependent kinase inhibitor p27 (Fig. 3c, and Supplementary Fig. 7f, g)³. Next we determined whether expression of the Id2 mutants that are resistant to APC/C^{Cdh1}-mediated degradation stimulated axonal elongation. CGN axons are identified on the basis of morphology, positive staining for Tau, and negative staining for MAP-2 (refs 3, 12). In the presence of green fluorescent protein (GFP) expression to mark transfected cells, expression of Id2-DBM and Id2 δ DB but not wild-type Id2 significantly increased axonal length in CGNs, cortical neurons and differentiated SK-N-SH cells (Fig. 3d, e, Supplementary Figs 7h and 8a, b, and data not shown). The same effect was evident in postnatal day 6 (P6) CGNs transfected with Id2-DBM and cultured on top of organotypic cerebellar slices from P9 rat pups³ (Supplementary Fig. 8c, d).

Having shown that, resembling Cdh1 silencing, the expression of undegradable Id2 leads to unrestrained axonal growth, we investigated whether restoring bHLH activity to neurons depleted of Cdh1 or expressing degradation-resistant Id2 rescues the deregulated axon growth. As expected, knockdown of Cdh1 in CGNs promoted axonal growth. Ectopic expression of the bHLH transcription factor E47 abolished the stimulation of axonal growth by Cdh1 knockdown and Id2-DBM (Fig. 3f–h). These results indicate that downstream targets of bHLH transcription factors mediate the axonal inhibitory functions and activation of these targets requires the control of Id protein accumulation by APC/C^{Cdh1}. To identify the relevant targets recruited by the Cdh1-Id-bHLH pathway for the control of axonogenesis, we infected SK-N-SH cells with an adenovirus expressing E47 (Ad-E47) and determined the gene expression profile. E47 induced the expression of genes coding for inhibitors of axonal growth including secreted molecules (Sema3F), ligands (Jagged-2) and receptors (Nogo receptor, Unc5A, Notch1) of multiple inhibitory and repellent signals for axons^{13–15} (Supplementary Fig. 9a). Prominent among them, the Nogo receptor is one of the best-known inhibitors of axonal growth on which all myelin-inhibitory signals (Nogo-A, myelin associated glycoprotein (MAG) and oligodendrocyte myelin glycoprotein) converge^{16,17}. We confirmed the microarray results with quantitative real time polymerase chain reaction from cells transduced with Ad-E47 and found that coinfection with Id2

adenovirus (Ad-Id2) prevented the E47-mediated induction of anti-axonal genes (Fig. 3i, and Supplementary Fig. 9b). The main intracellular transducer of axonal inhibitory signals driven by Nogo receptor is the small GTP-binding protein RhoA¹⁷. Ad-Id2 abolished the activation of RhoA by Nogo-A inhibitory peptide (Fig. 3j). To investigate whether bHLH transcription induces anti-axonal genes in non-neuronal cells, we generated SF210 glioma cells expressing conditionally active E47 fused to the ligand-binding domain of oestrogen receptor (E47-ER). Activation of bHLH transcription by 4-hydroxytamoxifen (4-OHT) increased Nogo receptor protein and the Val 1744-cleaved active fragment of Notch1 (Fig. 3k). Conversely, knockdown of Cdh1 decreased E47-mediated transcriptional activation of an E-box-luciferase plasmid and inhibited the basal and E47-induced expression of the anti-axonal signature and the Val 1744-cleaved intracellular fragment of Notch1 (Fig. 3c, and Supplementary Fig. 9c–e). Together, these results underscore the functional relevance of the Cdh1-Id-bHLH pathway for axonal growth.

To evaluate the significance of deregulated Id2 in neurons whose axonal growth had been arrested by the inhibitory myelin components that require Nogo receptor for signalling, we transfected CGNs plated on MAG-Fc inhibitory substrate with Id2-DBM or empty plasmid. Id2-DBM stimulated axonal growth in the absence of the MAG-Fc inhibitor and rescued axonal elongation in the presence of the myelin inhibitor. Conversely, MAG-Fc markedly inhibited axonal growth in neurons transfected with the empty plasmid (Fig. 4a, b). Next we determined whether Id2 stability differs between neurons that are actively growing axons and on reaching synaptogenesis when growth is terminated. Cortical neurons undergo axonal elongation *in vitro* that culminates with extensive synaptogenesis at DIV (days

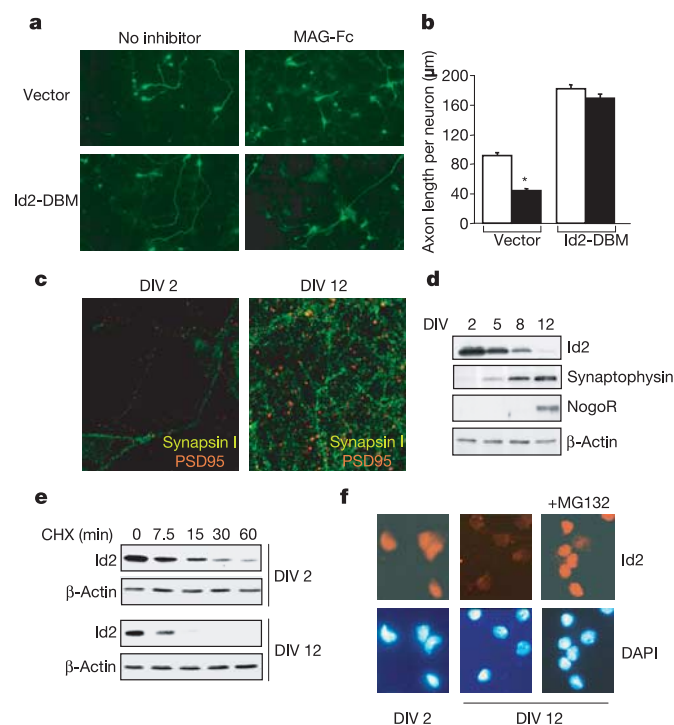


Figure 4 | Significance of APC/C^{Cdh1}-mediated degradation of Id2 for axonal growth. **a**, Axonal outgrowth of Id2-DBM and vector-transfected CGNs in the presence and absence of MAG-Fc. **b**, Quantification of axonal length from **a**. Open bars, no inhibitor; filled bars, MAG-Fc. Results are means $\pm$ s.e.m. ($n = 150$; asterisk, $P \leq 0.0001$). **c**, Co-staining for presynaptic synapsin-I (green) and postsynaptic PSD95 (red) in DIV 2 and DIV 12 cortical neurons. **d**, Id2 expression in cortical neurons undergoing synaptogenesis. NogoR, Nogo receptor. **e**, DIV 2 and DIV 12 cortical neurons treated with cycloheximide (CHX) for the indicated durations. **f**, Id2 (red) and nuclei (4,6-diamidino-2-phenylindole (DAPI), blue) staining in DIV 2 and DIV 12 cortical neurons in the presence and absence of MG132.

in vitro) 12 as shown by the accumulation of synaptophysin and maturation of the postsynaptic marker PSD-95 into large clusters juxtaposed to presynaptic synapsin-I puncta (Fig. 4c, d)¹⁸. The steady-state levels of Id2 decreased progressively after DIV 2 and became almost undetectable at DIV 12, when the Nogo receptor protein was markedly upregulated (Fig. 4d). The reduction of the Id2 steady state between DIV 2 and DIV 12 coincided with an increased rate of Id2 protein degradation (Fig. 4e). APC/C^{Cdh1} inhibition by silencing of Cdh1 led to an increase in Id1 and Id2 levels also in cortical neurons (Supplementary Fig. 10). APC/C^{Cdh1} activity is localized in the nucleus¹⁹. Consistent with this notion is the observation that MG132-mediated inhibition of the proteasome in DIV 12 neurons led to the accumulation of Id2 in the nucleus at levels comparable to those detected at DIV 2 (Fig. 4f).

We show that APC/C^{Cdh1} primes Id2 for degradation through a D box as a recognition site for the Cdh1 coactivator. The activity of Id proteins, which are invariably depleted in cells undergoing quiescence, corresponds to the consequences of inactivation of APC/C^{Cdh1} in postmitotic cells. Indeed, in a manner consistent with reactivation of the cell cycle in cells carrying inactive APC/C (ref. 20), deregulated expression of Id2 prevents cell cycle arrest by a wide range of antiproliferative signals^{21–23}. Under certain experimental conditions, ectopic Id2 is able to override the quiescent state and drive terminally differentiated cells back into the cell cycle²⁴. The observation that the most aggressive tumours frequently contain the largest amounts of Id proteins raises the possibility that inactivation of the control of Id protein stability by APC/C^{Cdh1} might also contribute to Id accumulation in cancer.

By setting the timing of Id protein degradation in the nervous system, APC/C^{Cdh1} initiates a chain of events that terminates with the activation of downstream targets of bHLH transcription factors. The dual activity of the brain APC/C^{Cdh1} as an inducer of the postmitotic state and an inhibitor of axonal growth can be viewed as an integrated process that restrains Id proteins during distinct stages of neural development and, as a consequence, sets the timing of bHLH-dependent transcription. The general implication of our findings is that neurons in active axonal growth should be viewed as relatively 'undifferentiated' compared with neurons that have reached their targets and are unable to resume neurite outgrowth. We suggest that, by disabling the APC/C^{Cdh1}–Id–bHLH axis, degradation-resistant forms of Id proteins might provide beneficial effects for axonal regeneration of damaged neurons.

METHODS

Transfections and siRNA. Cell transfection was performed with Lipofectamine 2000 (Invitrogen). CGNs, cortical neurons and SK-N-SH cells differentiated with retinoic acid and brain-derived neurotrophic factor²⁵ were transfected three times with 100 nM siRNA oligonucleotides and the Gene Silencer siRNA Transfection Reagent (Gene Therapy Systems)²⁶. siRNAs were Cdh1 Smart Pool from Dharmacon (5'-CGACAGAUCAUCAUCCAAA-3', 5'-CAACUGGAGC GUAACUUC-3', 5'-CGCCAGAGCUUCAGGACGA-3', 5'-AAGGGUCUCU UUACGUUU-3') and U6shCdh1 (5'-GGCUUCGUGCAGAUUGGGAA-3'). Lamin and firefly luciferase Smart Pools (Dharmacon) served as negative controls. **RhoA assay.** SK-N-SH cells were infected with Ad-vector or Ad-Id2 at a multiplicity of infection of 50. After 24 h, cells were stimulated by the addition of the Nogo-A peptide Nogo-P4 (8 µM; Alpha Diagnostic) for 30 min. GTP-RhoA was precipitated by using beads with GST–Rho-binding domain (RBD) of rhotekin, and detected with anti-RhoA antibodies in accordance with the manufacturer's guidelines (Cytoskeleton). Additional methods are described in Supplementary Information.

Received 27 March; accepted 12 May 2006.

Published online 28 June 2006.

1. Iavarone, A. & Lasorella, A. Id proteins in neural cancer. *Cancer Lett.* **204**, 189–196 (2004).
2. Perk, J., Iavarone, A. & Benezra, R. Id family of helix–loop–helix proteins in cancer. *Nature Rev. Cancer* **5**, 603–614 (2005).
3. Konishi, Y., Stegmüller, J., Matsuda, T., Bonni, S. & Bonni, A. Cdh1-APC controls

axonal growth and patterning in the mammalian brain. *Science* **303**, 1026–1030 (2004).

4. Peters, J. M. The anaphase-promoting complex: proteolysis in mitosis and beyond. *Mol. Cell* **9**, 931–943 (2002).
5. Stegmüller, J. & Bonni, A. Moving past proliferation: new roles for Cdh1-APC in postmitotic neurons. *Trends Neurosci.* **28**, 596–601 (2005).
6. Bounpheng, M. A., Dimas, J. J., Dodds, S. G. & Christy, B. A. Degradation of Id proteins by the ubiquitin–proteasome pathway. *FASEB J.* **13**, 2257–2264 (1999).
7. Fajerman, I., Schwartz, A. L. & Ciechanover, A. Degradation of the Id2 developmental regulator: targeting via N-terminal ubiquitination. *Biochem. Biophys. Res. Commun.* **314**, 505–512 (2004).
8. Wainwright, L. J., Lasorella, A. & Iavarone, A. Distinct mechanisms of cell cycle arrest control the decision between differentiation and senescence in human neuroblastoma cells. *Proc. Natl Acad. Sci. USA* **98**, 9396–9400 (2001).
9. Reimann, J. D. *et al.* Emi1 is a mitotic regulator that interacts with Cdc20 and inhibits the anaphase promoting complex. *Cell* **105**, 645–655 (2001).
10. Harper, J. W., Burton, J. L. & Solomon, M. J. The anaphase-promoting complex: it's not just for mitosis any more. *Genes Dev.* **16**, 2179–2206 (2002).
11. Gieffers, C., Peters, B. H., Kramer, E. R., Dotti, C. G. & Peters, J. M. Expression of the CDH1-associated form of the anaphase-promoting complex in postmitotic neurons. *Proc. Natl Acad. Sci. USA* **96**, 11317–11322 (1999).
12. Powell, S. K., Rivas, R. J., Rodriguez-Boulton, E. & Hatten, M. E. Development of polarity in cerebellar granule neurons. *J. Neurobiol.* **32**, 223–236 (1997).
13. Barallobre, M. J., Pascual, M., Del Rio, J. A. & Soriano, E. The Netrin family of guidance factors: emphasis on Netrin-1 signalling. *Brain Res. Brain Res. Rev.* **49**, 22–47 (2005).
14. Fiore, R. & Puschel, A. W. The function of semaphorins during nervous system development. *Front. Biosci.* **8**, s484–s499 (2003).
15. Sestan, N., Artavanis-Tsakonas, S. & Rakic, P. Contact-dependent inhibition of cortical neurite growth mediated by notch signaling. *Science* **286**, 741–746 (1999).
16. Schwab, M. E. Nogo and axon regeneration. *Curr. Opin. Neurobiol.* **14**, 118–124 (2004).
17. Spencer, T., Domeniconi, M., Cao, Z. & Filbin, M. T. New roles for old proteins in adult CNS axonal regeneration. *Curr. Opin. Neurobiol.* **13**, 133–139 (2003).
18. Lesuisse, C. & Martin, L. J. Long-term culture of mouse cortical neurons as a model for neuronal development, aging, and death. *J. Neurobiol.* **51**, 9–23 (2002).
19. Zhou, Y., Ching, Y. P., Chun, A. C. & Jin, D. Y. Nuclear localization of the cell cycle regulator CDH1 and its regulation by phosphorylation. *J. Biol. Chem.* **278**, 12530–12536 (2003).
20. Wirth, K. G. *et al.* Loss of the anaphase-promoting complex in quiescent cells causes unscheduled hepatocyte proliferation. *Genes Dev.* **18**, 88–98 (2004).
21. Baghdoyan, S. *et al.* Id2 reverses cell cycle arrest induced by γ-irradiation in human HaCaT keratinocytes. *J. Biol. Chem.* **280**, 15836–15841 (2005).
22. Kowanetz, M., Valcourt, U., Bergstrom, R., Heldin, C. H. & Moustakas, A. Id2 and Id3 define the potency of cell proliferation and differentiation responses to transforming growth factor β and bone morphogenetic protein. *Mol. Cell. Biol.* **24**, 4241–4254 (2004).
23. Lasorella, A. *et al.* Id2 is critical for cellular proliferation and is the oncogenic effector of N-myc in human neuroblastoma. *Cancer Res.* **62**, 301–306 (2002).
24. Chaudhary, J., Sadler-Riggelman, I., Ague, J. M. & Skinner, M. K. The helix–loop–helix inhibitor of differentiation (ID) proteins induce post-mitotic terminally differentiated sertoli cells to re-enter the cell cycle and proliferate. *Biol. Reprod.* **72**, 1205–1217 (2005).
25. Encinas, M. *et al.* Sequential treatment of SH-SY5Y cells with retinoic acid and brain-derived neurotrophic factor gives rise to fully differentiated, neurotrophic factor-dependent, human neuron-like cells. *J. Neurochem.* **75**, 991–1003 (2000).
26. Aarts, M. *et al.* A key role for TRPM7 channels in anoxic neuronal death. *Cell* **115**, 863–877 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Gygi and N. Sherman for the identification of Id2-associated proteins by mass spectrometry; P. K. Jackson for providing us with purified MBP-Emi1; and W. G. Kaelin and J. Lukas for the U2OS cells conditionally expressing Myc-Cdh1. This work was supported by grants from the National Institutes of Health to A.L., A.I., A.B. and M.P., from the Charlotte Geyer Foundation to A.I., and from the Christopher Reeve Paralysis Foundation to A.B. J.S. and D.G. are supported by Deutsche Forschungsgemeinschaft and Emerald Foundation grants, respectively.

Author Information The microarray data have been deposited in the ArrayExpress database (<http://www.ebi.ac.uk/arrayexpress/query/entry>) under accession number E-MEXP-413. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.I. (ai2102@columbia.edu).

Molecular architecture of axonemal microtubule doublets revealed by cryo-electron tomography

Haixin Sui¹ & Kenneth H. Downing¹

The axoneme, which forms the core of eukaryotic flagella and cilia, is one of the largest macromolecular machines, with a structure that is largely conserved from protists to mammals¹. Microtubule doublets are structural components of axonemes that contain a number of proteins besides tubulin, and are usually found in arrays of nine doublets arranged around two singlet microtubules. Coordinated sliding of adjacent doublets, which involves a host of other proteins in the axoneme, produces periodic beating movements of the axoneme. We have obtained a three-dimensional density map of intact microtubule doublets using cryo-electron tomography and image averaging. Our map, with a resolution of about 3 nm, provides insights into locations of particular proteins within the doublets and the structural features of the doublets that define their mechanical properties. We identify likely candidates for several of these non-tubulin components of the doublets. This work offers insight on how tubulin protofilaments and accessory proteins attach together to form the doublets and provides a structural basis for understanding doublet function in axonemes.

Microtubule doublets isolated from sea urchin (*Strongylocentrotus purpuratus*) sperm were studied by cryo-electron tomography. Groups of up to nine parallel doublets, apparently from a single axoneme, are often found in our frozen-hydrated samples, as shown in Fig. 1 and Supplementary Fig. S2. The protofilaments (PFs) and also the 4 nm tubulin monomer repeat along the PFs are already resolved in the unaveraged tomographic reconstructions (Figs 1 and 2a). As described in Methods, small three-dimensional (3D) volumes along each doublet were extracted from the tomograms and then aligned and averaged. Combining data from nine doublets produced an improved density map, as illustrated in Fig. 2b. Because of the limited angular range over which data are collected the map shows better separation between PF densities in the plane of the specimen than perpendicular to it.

Axonemes contain a number of proteins in addition to tubulin, generally with periodicities that are multiples of the 8 nm tubulin dimer repeat. The longest periodicity detectable in our data from the isolated doublets was 16 nm, so the final map was obtained by filtering the Fourier transform using the layer lines at multiples of 1/(16 nm). Figure 2c is the axial projection of the final density map and shows that the microtubule doublet consists of a complete singlet microtubule, the A-tubule containing 13 PFs, and an incomplete microtubule, the B-tubule, containing ten PFs.

To interpret the structural features of the doublets, we built a pseudo-atomic model of the tubulin component by docking the crystal structure of the α/β tubulin PF (ref. 2) into the 3D density map (Fig. 3a). Unlike singlet microtubules which are circular in cross-section (as in Fig. 1c), the A-tubules show a slight elliptical deformation with an elongation of about 8% in the axoneme's radial dimension. We introduced this distortion in a model of a 13-PF microtubule with a mean radius of 112 Å to construct the A-tubule. The B-tubule was built of 10 straight PFs with a mean radius of 129 Å,

which corresponds to the size of a 15-PF microtubule, with a similar distortion in PF positions. For convenience in further discussion, we number PFs as shown in Fig. 3a. To judge the quality of the docking we calculated a density from the atomic model by removing data from its Fourier transform corresponding to the missing data for the experimental map. This calculated density matched the experimental density very well (Fig. 3b), accounting for effects of the anisotropic resolution. The model density accounts for almost all of the experimental density, and the high degree of similarity between the two maps throughout most of the structure makes it straightforward to identify features that represent non-tubulin components of the doublet. A difference map (Figs 3c–f) highlights densities that are clearly not accounted for by tubulin.

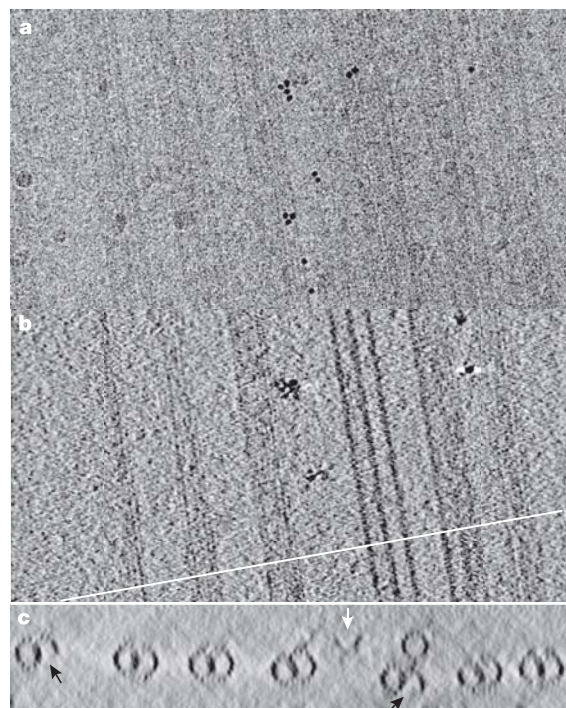


Figure 1 | Electron tomography of microtubule doublets. **a**, Doublets from a single axoneme are often found parallel to each other, as seen in the zero-tilt image from one tomographic series. **b**, Protofilaments are well resolved in the reconstructions, as shown in a 1-nm-thick section of this tomogram, roughly parallel to the doublets. **c**, A projection of a 26-nm-thick cross-section at the location indicated by the white line in Fig. 1b. The field includes five complete and two broken doublets (black arrows indicate missing PFs) as well as two singlet microtubules, one of which has partially disassembled (white arrow).

¹Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

Towards the outer side of the axoneme, PF B10 is in close contact with A4 and A5, apparently making a tight connection between the A- and B-tubules. In the present model the B10–A5 contact actually contains steric clashes, but the density does not provide sufficient resolution to modify the model reliably and resolve the clashes. Towards the inner side of the axoneme, PF B1 makes the closest approach to A1 but with a gap between the A- and B-tubules.

The region containing PFs A1–A4, referred to as the partition, covers the open part of the B-tubule and separates the hollow space of the two tubules. The PFs of this region can be isolated as a stable ribbon of three adjoining PFs^{3–6}. Our map shows how some of the other proteins bound in the partition region help to stabilize this ribbon.

The most distinctive density inside the A-tubule is next to the PFs of the partition, similar to features previously observed in cross-sections of plastic embedded doublets⁷. This density is resolved as a filamentous structure running mainly along PF A3 with projections extending out across several adjacent PFs. Figure 3d is a side view of this region showing three densities running across the partition. Two of these densities extend to PF A1, with an axial repeat of 8 nm, and the other runs across PFs A4 and A5 with a periodicity of 16 nm. These densities form a distinctive bridge-like structure in the longitudinal projection (Fig. 2c) which we term the partition bridge density.

The ribbon formed by the partition contains only a few proteins besides tubulin, including tektins A, B and C⁶. In transmission electron microscope (TEM) studies, a filament which has been shown to contain tektin is frequently seen extending from the end of the ribbon^{8–10} (see also Supplementary Fig. S3). Thus we surmise that the only continuous density in the tomogram associated with the ribbon PFs is tektin.

Tektins are predicted to share structural features with intermediate

filament proteins (IFPs)^{8,10–15}, forming heterodimers or homodimers with two globular heads and a coiled-coil structure extending in a tail. Cross-linking studies indicate that tektins A and B form a heterodimer¹⁶. The present density map is compatible with a number of models for how the tektins are arranged, but a likely interpretation is that the two domains pointing to the right in Fig. 3d correspond to parts of tektin A/B heterodimers, while the density pointing to the left is part of a tektin C homodimer. The continuous filament would then be composed of the head domains, possibly along with parts of the helical domains.

Also on the inside of the A-tubule, there are distinctive densities contacting PFs A7–A13. The density interacting with A10 and A11 protrudes into the lumen of the A-tubule. Another density between PF A12 and A13 projects to the outside of the A-tubule (Fig. 3a–c).

As shown in Fig. 3a, there is a small density projecting out from the doublet in the region of the outer junction between the A- and B-tubules. Because of the positioning of PF B10 in the present model, this feature does not appear in the difference density at the isosurface level used in Fig. 3c. However, if B10 is positioned to produce more realistic contacts with A4 and A5 one could interpret the density as being a small accessory protein that would stabilize the B10–A5 interaction. This protein could bind to both B10 and A5 in the region

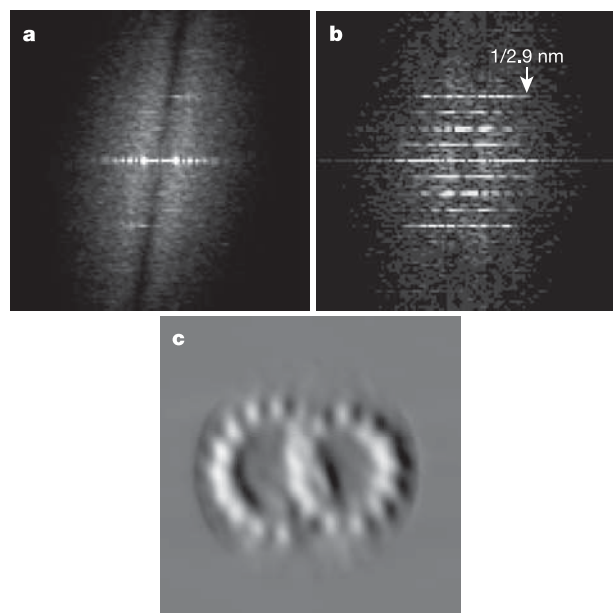


Figure 2 | Results of averaging tomographic data. **a**, Power spectrum of the projection of one doublet from a tomographic reconstruction. The doublet image was computationally straightened before computing the Fourier transform. The layer line at 4 nm is the most prominent. **b**, Power spectrum of the averaged tomogram from nine doublets, showing prominent layer lines at multiples of $1/(16 \text{ nm})$ and extending to a resolution better than 3 nm. **c**, Projection image of the final doublet density map along the longitudinal axis. The orientation was determined by comparison with other work in which doublets have been visualized in the context of the axoneme^{29,30}. The view is from the proximal end of the axoneme, and the upward direction corresponds to the outer direction of the axoneme.

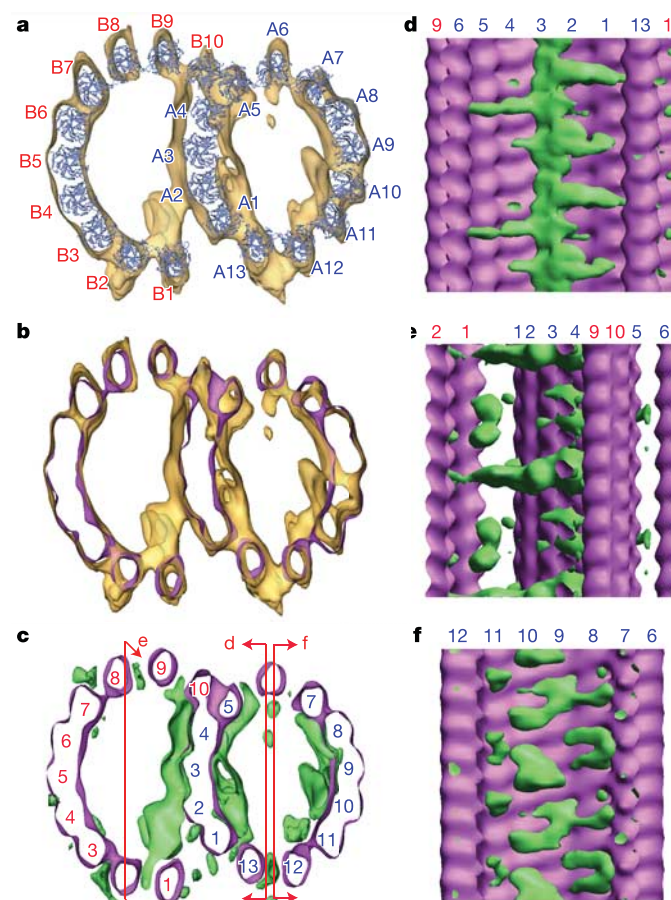


Figure 3 | Interpretation of the 3D density map. **a**, Axial view, seen from the proximal end, of the tomographic reconstruction that was filtered to enhance 16-nm spacings, viewed as an isosurface (yellow). The pseudo-atomic model (backbone only) fitted to the doublet density is shown in blue with labels assigned to the protofilaments. **b**, Projection view of the experimental density in yellow and density calculated from the pseudo-atomic model in purple. **c**, Same views as above with difference map shown in green and model PF density in purple. **d**, **e**, **f**, Side views of the difference map (green) and model density in purple. The orientations are indicated by the red lines in **c**.

occupied by other microtubule-associated proteins (MAPs), but its identity is still unclear.

Our 3D density map reveals a connection every 16 nm between the A and B-tubules towards the inner region (Fig. 3e), which we refer to as the linker density. The linker density connects PF A2 to B1 and extends to PF B2 on one side, with an extension that links PFs A2 and A3 on the outside of the A-tubule. The structure of this linker appears to provide a flexibility that would allow easier twisting of the doublet than if it were a rigid connection. There is another very different density along PF B1 between the linker densities. Some studies had suggested that there might be another PF of the B-tubule in this region, making a direct contact with the A-tubule, but our reconstruction shows that this is not the case.

The most likely candidates for the molecular components of the linker are identified as the polypeptides Sp77 and Sp83 in the isolated ribbons of sperm flagella⁶. Densities with a periodicity of 16 nm are observed at the edge of ribbons^{3,7} (see also Supplementary Fig. S3), where immunolabelling studies have identified both Sp77 and Sp83¹⁷, as well as Rib72, a *Chlamydomonas* homologue of Sp77 (refs 18, 19). Immunofluorescence microscopy has also shown that an anti-Sp77 antibody labels the nine doublets but not the central pair singlets, further supporting the hypothesis that Sp77 interacts with both the A- and B-tubules.

The linker density in the tomogram appears to represent a protein significantly larger than 77 kDa and thus probably contains other proteins. Because Sp77 and Sp83 are present in essentially the same amount, it is likely that Sp83 also contributes to the linker protein. Sp83 was found in both doublets and the central pair¹⁷, suggesting that its binding may be more similar to a conventional MAP on the A-tubule rather than directly linking to the B-tubule.

Within the axoneme, there are highly dynamic interactions between doublets arising from the inner and outer dyneins, as well as stable connections formed by nexin²⁰. Nexin has been localized in previous electron microscopy studies near the inner side of the doublets^{21,22}. In some of our tomograms, where the axonemes are less disrupted, we see a filamentous density that we interpret as nexin connecting PF B2 from one doublet to PFs A9–A11 on the adjacent doublet (Supplementary Fig. S2). Rib72 is also implicated in forming the inter-doublet linkages¹⁹ and may do so via interactions with nexin near PF B2.

Our interpretation of the arrangement of proteins found in the microtubule doublets along with associated proteins described elsewhere in the literature is summarized in Supplementary Fig. S1. The proteins in the partition bridge density, together with the linker density attached to the partition wall outside the A-tubule, clearly play a critical role in stabilizing the PFs near the partition. It appears that they may account for the elliptical distortion of the A-tubule, and they may also provide a substantial increase in resistance to bending of the doublet out of the plane. Other proteins of the A-tubule are located in regions where they could reinforce the tube and help to anchor proteins that bind on the outside. For example, there are distinctive densities inside the regions where the dynein heads are located. The largest feature, on PFs A10–A12, sits inside the attachment region for nexin and the dynein regulatory complex with which it is associated^{23–26}. The density we observe between PFs 12 and 13 is at the point where the radial spokes meet the doublets and may be part of their attachment. These external proteins have larger periodicities than the features we see in the doublets, and further work will be required to fully understand the significance of their interactions. The proteins that bind inside the A-tubule do so in a way very different from other MAPs, with the possible exception of tau²⁷, linking adjacent PFs in ways that seem to supplement the normal inter-PF interactions.

The configuration of the proteins that comprise the microtubule doublet appears to be designed to stabilize and maintain the protofilament architecture of the doublet as it undergoes the stresses involved in axoneme motion and also to favour bending in the

direction that corresponds to twisting of the axoneme. Our structure of the microtubule doublet provides insights into several novel tubulin–protein interactions and the functions doublets perform in axonemes, and will serve as a greatly improved basis for quantitative modelling of mechanical properties.

METHODS

Sperm flagella axonemes were prepared from sea urchin following the protocol in ref. 28. For the PF ribbon samples, material was treated with 0.7% Sarkosyl in 10 mM Tris-HCl (pH 7.8) at 4 °C for about 6 h and then centrifuged at 100,000 g for 30 min. The pellet was collected and resuspended in the above Tris buffer without Sarkosyl.

The concentrated doublet sample was resuspended in water containing 5 or 10 nm colloidal gold particles (Ted Pella), and applied to a grid covered with a glow-discharged holey carbon film. Grids were blotted, plunge-frozen in liquid ethane and transferred under liquid nitrogen to the cryo-tilt holder of a JEOL 3100FEF electron microscope. The microscope, operating at 300 kV, was equipped with an Omega energy filter set to a slit width of 25 eV. Microtubule doublets nearly parallel to the tilt axis were selected for data collection. Tomographic data were collected under low-dose conditions with accumulated dose of about 6,500 electrons per nm². Single-axis tilt series, including 33–46 images with a tilt increment of 3–4°, were recorded at a defocus of 3.5 to 5 µm on a 2,048 × 2,048 charge-coupled device camera (Gatan). The final pixel size of the images was about 0.53 nm.

Three-dimensional segment volumes along each doublet microtubule were extracted from the tomographic reconstructions. The volumes were aligned and averaged as described in Supplementary Information.

For modelling, the atom coordinates of the PF models for the A- and B-tubules were separately calculated as described in the text using the backbone atoms from the tubulin crystal structure, Protein Data Bank PDB ID 1JFF, and manually shifted into the density as a whole.

Received 6 February; accepted 21 April 2006.

Published online 31 May 2006.

- Kohl, L. & Bastin, P. The flagellum of trypanosomes. *Int. Rev. Cytol.* **244**, 227–285 (2005).
- Nogales, E., Wolf, S. G. & Downing, K. H. Structure of the $\alpha\beta$ tubulin dimer by electron crystallography. *Nature* **391**, 199–203 (1998).
- Witman, G. B., Carlson, K., Berliner, J. & Rosenbaum, J. L. *Chlamydomonas* flagella. I. Isolation and electrophoretic analysis of microtubules, matrix, membranes, and mastigonemes. *J. Cell Biol.* **54**, 507–539 (1972).
- Witman, G. B., Carlson, K. & Rosenbaum, J. L. *Chlamydomonas* flagella. II. The distribution of tubulins 1 and 2 in the outer doublet microtubules. *J. Cell Biol.* **54**, 540–555 (1972).
- Meza, I., Huang, B. & Bryan, J. Chemical heterogeneity of protofilaments forming the outer doublets from sea urchin flagella. *Exp. Cell Res.* **74**, 535–540 (1972).
- Linck, R. W. & Norrander, J. M. Protofilament ribbon compartments of ciliary and flagellar microtubules. *Protist* **154**, 299–311 (2003).
- Linck, R. W. Flagellar doublet microtubules: fractionation of minor components and α -tubulin from specific regions of the A-tubule. *J. Cell Sci.* **20**, 405–439 (1976).
- Linck, R. W. & Langevin, G. L. Structure and chemical composition of insoluble filamentous components of sperm flagellar microtubules. *J. Cell Sci.* **58**, 1–22 (1982).
- Linck, R. W., Amos, L. A. & Amos, W. B. Localization of tektin filaments in microtubules of sea urchin sperm flagella by immunoelectron microscopy. *J. Cell Biol.* **100**, 126–135 (1985).
- Linck, R. W. & Stephens, R. E. Biochemical characterization of tektins from sperm flagellar doublet microtubules. *J. Cell Biol.* **104**, 1069–1075 (1987).
- Chang, X. J. & Piperno, G. Cross-reactivity of antibodies specific for flagellar tektin and intermediate filament subunits. *J. Cell Biol.* **104**, 1563–1568 (1987).
- Linck, R. W., Albertini, D. F., Kenney, D. M. & Langevin, G. L. Tektin filaments: chemically unique filaments of sperm flagellar microtubules. *Cell Motility Suppl.* **1**, 127–132 (1982).
- Chen, R., Perrone, C. A., Amos, L. A. & Linck, R. W. Tektin B1 from ciliary microtubules: primary structure as deduced from the cDNA sequence and comparison with tektin A1. *J. Cell Sci.* **106**, 909–918 (1993).
- Norrander, J. M., Amos, L. A. & Linck, R. W. Primary structure of tektin A1: comparison with intermediate-filament proteins and a model for its association with tubulin. *Proc. Natl Acad. Sci. USA* **89**, 8567–8571 (1992).
- Steffen, W. & Linck, R. W. Relationship between tektins and intermediate filament proteins: an immunological study. *Cell Motil. Cytoskel.* **14**, 359–371 (1989).
- Pirner, M. A. & Linck, R. W. Tektins are heterodimeric polymers in flagellar microtubules with axial periodicities matching the tubulin lattice. *J. Biol. Chem.* **269**, 31800–31806 (1994).

17. Hinchcliffe, E. H. & Linck, R. W. Two proteins isolated from sea urchin sperm flagella: structural components common to the stable microtubules of axonemes and centrioles. *J. Cell Sci.* **111**, 585–595 (1998).
18. Patel-King, R. S., Benashski, S. E. & King, S. M. A bipartite Ca^{2+} -regulated nucleoside-diphosphate kinase system within the *Chlamydomonas* flagellum. The regulatory subunit p72. *J. Biol. Chem.* **277**, 34271–34279 (2002).
19. Ikeda, K. *et al.* Rib72, a conserved protein associated with the ribbon compartment of flagellar A-microtubules and potentially involved in the linkage between outer doublet microtubules. *J. Biol. Chem.* **278**, 7725–7734 (2003).
20. Lindemann, C. B. Testing the geometric clutch hypothesis. *Biol. Cell.* **96**, 681–690 (2004).
21. Bozkurt, H. H. & Woolley, D. M. Morphology of nexin links in relation to interdoublet sliding in the sperm flagellum. *Cell Motil. Cytoskel.* **24**, 109–118 (1993).
22. Woolley, D. M. Studies on the eel sperm flagellum. I. The structure of the inner dynein arm complex. *J. Cell Sci.* **110**, 85–94 (1997).
23. Huang, B., Ramanis, Z. & Luck, D. J. Suppressor mutations in *Chlamydomonas* reveal a regulatory mechanism for flagellar function. *Cell* **28**, 115–124 (1982).
24. Gardner, L. C., O'Toole, E., Perrone, C. A., Giddings, T. & Porter, M. E. Components of a "dynein regulatory complex" are located at the junction between the radial spokes and the dynein arms in *Chlamydomonas* flagella. *J. Cell Biol.* **127**, 1311–1325 (1994).
25. Piperno, G., Mead, K., LeDizet, M. & Moscatelli, A. Mutations in the "dynein regulatory complex" alter the ATP-insensitive binding sites for inner arm dyneins in *Chlamydomonas* axonemes. *J. Cell Biol.* **125**, 1109–1117 (1994).
26. Piperno, G., Mead, K. & Shestak, W. The inner dynein arms I2 interact with a "dynein regulatory complex" in *Chlamydomonas* flagella. *J. Cell Biol.* **118**, 1455–1463 (1992).
27. Kar, S., Fan, J., Smith, M. J., Goedert, M. & Amos, L. A. Repeat motifs of tau bind to the insides of microtubules in the absence of taxol. *EMBO J.* **22**, 70–77 (2003).
28. Waterman-Storer, C. M. in *Current Protocols in Cell Biology* (eds Bonifacino, J. S., Dasso, M., Harford, J. B., Lippincott-Schwartz, J. & Yamada, K. M.) 13.1.6–13.1.7 (John Wiley, New York, 1998).
29. Hoops, H. J. & Witman, G. B. Outer doublet heterogeneity reveals structural polarity related to beat direction in *Chlamydomonas* flagella. *J. Cell Biol.* **97**, 902–908 (1983).
30. Mastronarde, D. N., O'Toole, E. T., McDonald, K. L., McIntosh, J. R. & Porter, M. E. Arrangement of inner dynein arms in wild-type and mutant flagella of *Chlamydomonas*. *J. Cell Biol.* **118**, 1145–1162 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Killilea for sea urchin collection, K. Gull for discussions on protofilament numbering, H. Li, B. Rockel and D. Typke for discussions and help with image processing, B. Heymann and C. Yang for discussions and help with BSOFT library usage, and R. Glaeser and M. Auer for suggestions and encouragement. This work is supported by NIH grants and by the US Department of Energy.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.H.D. (khdowning@lbl.gov).

naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

It seemed like a curveball question. The *Naturejobs*-sponsored panel at the EuroScience Open Forum in Munich last week had been discussing scientific opportunities outside academia for about two hours when an audience member asked whether female scientists should have children early in their career, or wait until after they are well established. The query seemed left-field enough, given the topic being debated, but the gender of the questioner surprised both the audience and the panellists: he was male.

Perhaps the issue arose because four of the seven panel members were women — and some of them had mentioned having children when they introduced their talk. A couple of the male panellists followed their lead, noting that family issues affect women disproportionately — after all, there are plenty of data to show that compared with men, women's scientific careers are more hindered by starting a family. So the issue ought to be considered by scientists of both sexes.

In response to the question, the panellists concurred that timing child-bearing with defending your dissertation seems logical; but they emphasized that cold-blooded rationale isn't the best consideration for such decisions. Several spoke of how they managed to advance their careers even though they had started a family. One woman went part-time after having her baby and has now increased her hours to 80% of her full-time post. The set-up has worked well, she said, although there is the slight issue that her academic husband works in another country. Another female panellist has her own patent-law business and so can build flexibility into her own schedule.

The ultimate solution to dealing with the issue of when to have children requires supportive partners and institutions, but the decision is ultimately down to the individual. Just as it's hard to plot the entire course of one's career at the beginning of graduate school, it's equally hard to lay out one's entire family plan: marriage and honeymoon between grant cycles; children timed to appear right after tenure; and death — but only after one's had time to properly enjoy one's Nobel.

Paul Smaglik, *Naturejobs* editor

CONTACTS

Publisher: Ben Crowe

Editor: Paul Smaglik

Assistant Editor: Gene Russo

European Head Office, London

The Macmillan Building,

4 Crinan Street

London N1 9XW, UK

Tel: +44 (0) 20 7843 4961

Fax: +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

European Sales Manager:

Andy Douglas (4975)

Natureevents: Matt Betteridge (4015)

UK/RoW/Ireland:

Nils Moeller (4953)

Scandinavia/Spain/Portugal:

Claudia Paulsen Young (4973)

France/Switzerland/Belgium/Italy:

Maria Ceraolo (4994)

Germany/Austria/The Netherlands:

Reya Silao (4970)

Business Development Manager:

Amelie Pequignot (4974)

Advertising Production Manager:

Stephen Russell

To send materials use London

address above.

Tel: +44 (0) 20 7843 4816

Fax: +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Catherine Alexander

European Satellite Office

Germany:

Patrick Phelan

Tel: +49 89 54 90 57 11

Fax: +49 89 54 90 57 20

e-mail: p.phelan@nature.com

US Head Office, New York

75 Varick Street,

9th Floor, New York,

NY 10013-1917

Tel: +1 800 989 7718

Fax: +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

Chiyoda Building,

2-37 Ichigayatamachi,

Shinjuku-ku,

Tokyo 162-0843

Tel: +81 3 3267 8751

Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:

Ayako Watanabe

e-mail: a.watanabe@natureasia.com



IMAGES.COM/CORBIS

TRIAL BLAZERS

The drug industry may be going through lean times, as new candidates have to clear ever-higher safety hurdles. But this gives scientists who can steer a drug through clinical trials a head start in the job market, says **Hannah Hoag**.

When the draft sequence of the human genome appeared in 2000, pharmaceutical and biotechnology executives anticipated an increase in drugs approved for the marketplace. Instead, the reverse happened; the number of new drug approvals by the US Food and Drug Administration declined until 2004. And although 35 new drugs were approved that year, only 20 new drugs reached the US market in 2005. Drugs rarely sail through clinical trials, the safety and efficacy tests that determine whether a drug will reach the market and stay there (see 'Clinical trials step by step'). Before 2000, drugs entering the first stage of testing on humans, known as phase I trials, had a 14% chance of reaching the market; today that figure is a mere 7%.

Because companies are having such a difficult time getting drugs approved, scientists adept at planning, running and managing clinical trials will find themselves awash with opportunities. The global growth of contract research organizations (CROs), which help

drug companies with preclinical work and clinical trials, combined with increased scrutiny of clinical trials, adds up to ample job opportunities. "There is high demand, but a very constrained labour pool," says Chris Jock, vice-president and general manager for global operations at Kelly Scientific Resources, Troy, Michigan. Increased emphasis on regulation, pressure to produce data faster, the opening of the global market and a shift towards generics has put unusual pressure on the drug industry. This supply-demand imbalance is unlikely to vanish in the near future, says Jock.

Among the many professionals who participate in clinical trials (see 'Stats and stories'), two key positions provide most support: the clinical research associates (CRAs) and the more senior clinical research scientists. "We recruit heavily for both positions," says Steven Smith, director of human resources and development at Novartis in East Hanover, New Jersey. Both generally demand a background in life sciences and excellent communication skills.

The CRA, or 'field monitor', checks the clinical study from start to finish — identifying the doctors who can find volunteers, educating them about the study, doing the paperwork and ensuring that supplies are on hand. CRAs should be willing to travel as a lot of time is spent visiting trial sites. "They are doing the blocking and the tackling of clinical-trial execution," says Chip Clark of Vanda Pharmaceuticals, a small firm in Rockville, Maryland, that applies genomic and genetic tools to undeveloped drug compounds to try to find unexpected indications or more appropriate patient subpopulations.

Once the trial begins, the field monitor checks the

Ample opportunities:
Chris Jock says there is high demand for clinical-research scientists.



data for significant adverse events, notation errors and omissions, before passing them along to the data-management group. More senior CRAs may also write protocols, abstracts and manuscripts.

For the clinical research scientist, the job focus lies in trial design. They assist doctors and coordinators with the trial, provide training, aid in data collection and may work with the regulatory-affairs team. Doctoral degrees in the biological sciences are common among this group. "They are the internal contact for the investigators, there to assist if there are questions about the protocol, the dosing or payments. They are the heart," says Lisa Jean-Louis, associate director of clinical research at Novartis.

The right skills

Three to five years' clinical or industry experience may open up more senior positions. Applicants with less experience should maximize their networking, or take extra courses or certificate programmes through industry associations, such as the Association of Clinical Research Organizations or the Society of Clinical Research Associates, says Leslie Eggleston, manager of corporate staffing at Otsuka Maryland Research Institute in Rockville, Maryland. An internship at a drug company can provide young scientists with a foot in the door.

Project management is the other main track. "Project managers often have a life-science background, but they are focused on processes and leading teams," says Sam Bidwell, senior vice-president of human resources and global product development services at Quintiles, a CRO based in Research Triangle Park, North Carolina. The project manager's responsibilities include site selection, feasibility studies, customer relations, budget planning and data management. "A business background is preferred, but the science background will facilitate their understanding of the processes," says Eggleston, adding that these positions are often difficult to fill.

The profile of an ideal candidate depends largely on the sector. Although pharmaceutical companies, biotechs and CROs all rely on professionals to take a drug compound through clinical trials, they vary in the number of positions available, the responsibilities and the experience required. Candidates stand out if they have expertise in a specific therapeutic area, especially if the study has already started. "It means only getting

STATS AND STORIES

Biostatisticians are a key piece of the clinical-trial puzzle, although they often go unnamed. "We generate reams and reams of data and we have to be able to interpret them," says Evan Loh, vice-president of Clinical Research and Development at Wyeth Pharmaceuticals in Madison, New Jersey. But the job goes far beyond sorting data. "When doing hypothesis-driven research, you have to prospectively design your hypothesis and your analysis tools in order to interpret the data in a non-biased way," says Loh. A background in a biological discipline can be a plus, but is rarely required.

It takes about 12 years for an experimental drug to gain approval from the Food and Drug Administration — six of these years are spent in clinical trials. Without talented medical writers, all this work is wasted. These writers plan, write and edit study reports, manuscripts and other related clinical documents. They frequently hold advanced scientific or medical degrees and have some academic or industry writing experience. **H.H.**

them acclimated to the Novartis ways, not having to train them in diabetes," says Jean-Louis.

Pharmaceutical companies may have extensive in-house clinical-trials staff, or maintain a smaller team of clinical research scientists and project managers to work with the CRO. Smaller companies can offer a career with greater visibility, independence and responsibility, but they may have more stringent hiring requirements. "We are being very careful about who we hire because we are just 35 people," says Clark. "One bad hire would stick out like a sore thumb. We really need to get the first 50 or 60 people right." Small biotechs often hire people from drug firms.

Well connected

Most of the growth in the job market lies with CROs. Worldwide, the CRO market is worth about \$15 billion this year, and its revenue is increasing at an annual rate of about 15%. CROs are known for their speed and efficiency; they can complete a clinical trial in two-thirds the time a drug company can, shaving months off the process and offering \$120 million to \$150 million in increased revenue per drug. Small or medium-size drug firms and biotech companies know that the farther the compound moves through the research cycle, the more money they can raise. "Of the top 30 best-selling drugs, we've touched every one," says Bidwell.

The CROs have global reach, recruiting trial subjects at speeds often unmatched by drug companies (see *Nature* **441**, 22–23; 2006). According to the Center for the Study of Drug Development at Tufts University in Boston, Massachusetts, nearly one-third of US drug trials are conducted abroad. Smith estimates that 60% of Novartis's trials are run overseas, in Europe, Latin America and more recently India and China. Eggleston notes that it can be difficult to find US candidates with the necessary foreign experience. Bidwell says that Quintiles has employees in 53 countries around the world and hires many local monitors.

Altogether, the job market in clinical trials is so hot that a well-qualified candidate can choose among thousands of positions available globally, and has a fair chance of landing one within a month. ■

Hannah Hoag is a freelance writer based in Montreal, Canada.

CLINICAL TRIALS STEP BY STEP

Phase I trials evaluate the safety, tolerability, pharmacokinetics and pharmacodynamics of a drug in 20 to 80 volunteers. Most phase I clinical trials are conducted in an in-patient clinical setting on healthy volunteers.

Phase II trials measure the drug's efficacy in 100 to 300 volunteers and patients.

Phase III trials comprise 1,000 to 3,000 volunteers, or even more. The trial is often double-blind, randomized and controlled.

Phase IV trials involve post-launch safety and surveillance of a drug, to detect rare or long-term adverse effects.

H.H.



MOVERS

Matthias Kleiner, president, DFG, Bonn, Germany



2004-06: Head of the Institute of Forming Technology and Lightweight Construction, University of Dortmund, Germany.

1998-2006: Professor of forming technology, University of Dortmund, Germany.

1994-98: Professor, University of Cottbus, Germany.

When Matthias Kleiner leaves academia in January to become president of the DFG, Germany's main research funding agency, he will be the first engineer to take on the country's most influential research-management position. He will face challenges relating to financing research, expanding university research and dealing with Germany's restrictions on stem-cell science. "I am entering new territory," he says, "but this is what a scientist is used to."

Kleiner's research field — which focuses on engineering items such as light-weight metals for high-speed trains — rarely makes headline news. As DFG president, Kleiner will oversee the distribution of grants and awards worth €1.3 billion (US\$1.6 billion) per year, funding everything from quantum physics to the social sciences.

After college, Kleiner chose science over acting, and later rejected several industry job offers to stay in academia. In 1994, four years after German unification, he became full professor at the University of Cottbus in eastern Germany. For a young professor in difficult scientific environs, setting up a new department for engineering sciences was challenging. The lack of industry collaborations in the east was a factor in his later return to western Germany.

A big part of Kleiner's new job will be managing a €1.9-billion 'excellence initiative' aimed at creating top universities and establishing several new graduate schools. He is keen to improve career opportunities for scientists in Germany and to encourage talented young researchers. "Develop an appetite for science," he tells them. "Be curious and look to the right and left alongside the path you choose."

Politically, Kleiner's most delicate task will be addressing the topic of human embryonic stem-cell research. Restrictive laws have led some scientists to leave Germany. Before making any judgement, Kleiner says he needs a better understanding of the field and the ethics. The technology can be used responsibly, he says, but regulations will be required.

Noting that inadequate technology transfer has been identified as partially responsible for Germany's sluggish economy, Kleiner endorses more collaboration between industry and academia. Industry, he says, should more frequently be involved in DFG-funded research projects in order to facilitate downstream applications.

"It is about time for an engineer to head the DFG," says Hubert Markl, a zoologist and former president of the DFG. "I know they always wanted one, but apparently most of these guys are too busy with other things." ■
Friederike Siegel

RECRUITERS & ACADEMIA

Where are the physician-scientists?

Recently, a resident of mine was trying to decide whether to pursue a career in academic medicine. "Why would I want to be a physician-scientist?" he asked me. "Isn't it difficult to get funding and aren't the chances of success low?" These concerns are legitimate. I have seen many colleagues begin their careers as academic physicians only to move on to non-academic jobs.

To train a future generation of scientists that can bridge the gap between the laboratory and the patient, we need more programmes like the 'Physician-Scientist Early Career Award' run by the Howard Hughes Medical Institute (HHMI) and the KO8 'Mentored Clinical Scientist Award' from the National Institutes of Health (NIH). By providing extra funding, grants such as these free up time for laboratory research, seminars, courses, journal clubs, and national and international meetings.

Since first applying for an HHMI physician-scientist fellowship almost ten years ago, I've believed that the hybrid position offers unique and rewarding opportunities. As clinicians, we are able to provide compassionate care for patients. As scientists, we can investigate the pathological basis of disease and discover new treatments for devastating illnesses.

In my current job as director of a brain-tumour programme, I maintain a strong presence in both the laboratory and the clinic. The two responsibilities are truly complementary. The patients in my clinical programme motivate me to research the mechanisms of brain-tumour development, and my laboratory research addresses basic questions about the human brain and brain tumours. Eventually, I hope to help translate this knowledge into improved therapies for brain-tumour patients. And, with the support of the NIH and the HHMI, I hope to have the opportunities and resources to investigate and develop new ways to fight this disease.

So far, however, we have not made significant strides towards innovative treatment strategies. As my resident and I began another day studying the brain, I emphasized that it is up to us to change the harsh decree that 'nothing can be done' to help these patients. Despite having concerns about future funding for physician-scientists, he said he would think seriously about this path. I hope he decides to join me. ■

Alfredo Quiñones-Hinojosa is director of the brain-tumour surgery programme at Johns Hopkins Bayview Medical Center.

GRADUATE JOURNAL

Lab makeover

As a high-school student I had a vision of graduate school, probably fuelled by watching too many French art-house films. I imagined erudite academics sitting in coffee shops drinking espresso. Now, at the end of my graduate career, I'm finally living my dream. Except that, in reality, multiple cups of coffee make my hands shake and the questions I'm working on are so abstruse that most of my conversations are with myself.

Before my coffee-shop phase, I was in what I call my worker-bee period — spending most of my time collecting data in the lab. This reflects the intrinsic contradiction of becoming a successful lab scientist. On the one hand you need to doggedly repeat experiments with robotic precision, and on the other you need to think creatively and freely about your science.

Returning to the lab after another spell at the coffee shop, I had an idea for speeding up the development of future graduate students: combine the two phases of graduate school. To do so, I suggest changing the ambience of our sterile labs. Add some light jazz, a few paintings by local artists, and berets. This, as my dream demonstrates, will inspire graduate students to greater intellectual heights even while they're methodically collecting data. If only our safety office would let me install an espresso machine next to my lab bench... ■

Milan de Vries is a molecular-biology graduate student at the Massachusetts Institute of Technology.

Golden year

The gift of memories.

Igor Teper

It was only when Will found himself among the poplars, maples and oaks of a grove in Aventine Habitat's eastern park that he figured out why he'd felt so antsy all morning. Why he was so restless that he'd decided to defy the aches and lethargy that were now his constant companions, to leave his room for the first time in weeks and go for a long walk through the habitat. Only when his legs, as if hearkening back to his younger days, had brought him to stand among those well-remembered trees did he realize that this would have been the day of his fiftieth wedding anniversary.

Alice had modelled the grove on the clearing in the Maine woods where he had proposed to her, during a hike on a brisk October afternoon more than half a century before. The inspiration for the grove's design had been their secret, and they had come there for every anniversary after they moved to the Moon — 36 anniversaries in all. The memory of those visits was a time-lapse film of the trees' growth, from seedlings to saplings to giants that now almost completely blocked the artificial daylight streaming down from the underside of the habitat's dome.

Alice's horticultures were all over the habitat. Trees moulded by both their genes and the ecomaintenance programs that controlled the temperature, humidity, lighting and soil-nutrient content so that they bent and intertwined as they grew to form domed enclosures, gazebos and arches, or twisted into helices and braids, or lifted and drooped their branches into fountains and waterfalls of leaves.

But it was the grove, simpler and quieter than her other creations, and more private, that had been Alice and Will's special place. For a long time after Alice received the offer from Aventine Habitat, Will had resisted the idea of moving to the Moon. He had not wanted to abandon the life they'd started to build in the place where he'd lived all his life. In the end, he acquiesced, and to thank him for giving her the

opportunity to shape the greenery of an entire city, Alice had used her singular talents to recreate a piece of the New England he missed so intensely.

As he saw the grove take shape over the years, Will had been amazed at the faithfulness of the reproduction. It not only looked but, on a visceral level, felt like the spot where he'd pretended to trip and met Alice's concern with a ring and a question.



The only difference was that, when he'd proposed, the trees had been alight with the most brilliant golds and crimsons of autumn, whereas the leaves in the seasonless habitat were always green. He'd once pointed this out to Alice, and she'd turned silent and looked away. He never mentioned it again.

Will had not returned to the grove since Alice's fatal illness, six years earlier, and it was now more like his memory than ever, a living bridge through time and space. It was as if Alice herself were reaching to him from the past, so close he could touch her.

He pressed his hand to the trunk of one of the largest maples, and was struck by

how much the bark, grey and brown, with light streaks and dark spots, looked like the age-mottled skin on the back of his hand. It occurred to him that even as the trees had grown into what they'd been meant to be, he had decayed far from the man Alice had agreed to marry.

He closed his eyes; the bark was warm and leathery beneath his fingers. A cool breeze caressed the back of his neck, and the ground seemed to shift beneath his feet. Leaves rustled overhead, and the air was filled with a scent so familiar it was unrecognizable. And with that scent, the memory of Alice looking down at him as he knelt before her that day in Maine flooded his senses, unbearably intense. Will opened his eyes, and did not believe them.

Gold, bronze, copper, amber, carnelians and rubies — the trees around him blazed with the most majestic display of autumn colours he'd ever seen, a swirling kaleidoscope that glowed and burned and sparkled. Face raised, mouth open in astonishment, he stepped back from the maple and slowly turned, trying to take in the entire scene, every tree, every branch, every leaf.

He could not begin to imagine what miracles of genetics and biochemistry had made the trees' transformation possible at his touch, but, for their golden wedding anniversary, Alice had given him autumn. She had prepared it all those decades ago, and had never, not even when she lay dying, given

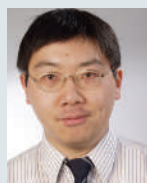
him any hints of what she'd planned. Still, some part of him must have known, for he had come to receive his gift, or perhaps she had known him so well that she'd foreseen his coming.

Will's heart, and his throat, and his thoughts seized up then; he missed Alice more acutely than he ever had, and he loved her more deeply than ever. Tears came pouring from within him, hot tears that watered the soil where Alice had planted her love, and, as he cried, leaves began to fall around him, like tears of gold. ■

Igor Teper teaches old atoms new tricks at temperatures just above absolute zero. He also writes stories, occasionally.

JACEY

Abstractions



FIRST AUTHOR

When the Huygens spacecraft descended into the atmosphere of Saturn's largest moon, Titan, in January 2005, the data it recorded suggested it had encountered a fine

mist of rain. But this was no ordinary rain — it was a methane mist. Some researchers thought methane condensation to be unlikely in Titan's frigid environment (its surface temperature is -180°C). However, according to the data collected by Huygens, which allowed an international team of scientists to calculate a profile of Titan's relative humidity, there is condensation on Titan. Tetsuya Tokano, a geophysicist at the University of Cologne, Germany, and his colleagues used Huygens' data to characterize Titan's clouds (see page 432). He tells *Nature* about methane rain on a distant moon.

Had you always had a fascination with Titan?

Clouds on Earth were the subject of my master's thesis. I then moved on to study Titan, developing climate and atmospheric-circulation models. I was interested to know whether, if rain arrives at the moon's surface, there would be a hydrological cycle. As it turns out, there is.

How often does it rain on Titan?

We believe that it was raining during the descent. In contrast to sporadic clouds observed near the South Pole, this rainfall is permanent for at least the season, which lasts several years on Titan. There are 30 Earth years to every one Titan year.

What does methane drizzle suggest about the geological features of Titan?

The rivers on Titan's surface were probably not caused by this drizzle because the rainfall is too weak. But the presence of drizzle does not rule out the occurrence of thunderstorms and heavy rain storms at other times.

How do your results compare with those of Hueso and Sánchez-Lavega (page 428)?

They believe that rainstorms are rare but violent, whereas we feel that rain is ubiquitous but weak. However, I think both can occur on Titan, although not simultaneously.

Has such drizzle been documented on other planets and moons?

Methane clouds have been identified on some other planets, but, so far, no actual rainfall has been observed.

Do your results have any relevance for the hydrological cycle on Earth?

This paper would definitely be of interest to terrestrial meteorologists. The comparison of methane and water condensation provides a better idea of how condensation works. ■

MAKING THE PAPER

Antonio Iavarone

How a simple protein sent researchers on a long journey.

Sometimes, finding out how proteins interact with each other, trigger other reactions and affect development is more complicated than the biochemical process itself. Such was the case for two cancer biologists from Columbia University, New York. While Antonio Iavarone and his colleague Anna Lasorella were exploring the role of a family of proteins in brain cancer, their research led them from gene expression to basic cell-division processes, and on to neurobiology. During the course of this journey, which is detailed on page 471, they collaborated with two separate groups, eventually learning that a family of proteins essential to the basic processes of cell proliferation and cancer also exerts influence beyond cell division in nervous-system development.

Iavarone and Lasorella had been studying a family of proteins to find out whether they were turned off or on during brain development and in brain cancer. These proteins, particularly one known as Id2, were already known to prevent stem cells from differentiating into adult cells. But the experimental data showed a paradox. Id2, which had been thought to be an inhibitor, actually seemed to be stimulating the development of axons — slim fibres that carry signals from nerve to nerve. "We couldn't figure out what it meant," says Iavarone. "We had the result for a year without really knowing how to follow it up."

Meanwhile, Lasorella and Iavarone had been using proteomics to identify the partners of Id2 in neuronal cells. They discovered that Id2 interacts with the anaphase promoting complex (APC), a cluster of proteins that works by priming other proteins for degradation and is key to the cell-division process of mitosis. "We looked for how the proteins in the APC complex talked to each other, how they attached to



each other and which proteins in APC attached to Id2," says Iavarone.

With the help of a cancer-biology group at New York University (NYU) that had expertise in looking at biochemical processes triggered by APC, Lasorella and Iavarone found that when APC links up with Id2, the complex then degrades the protein. But when mutations that prevented binding to APC were introduced in Id2, the protein became resistant to degradation and accumulated inside the cells at very high levels.

Meanwhile, a paper from a group at Harvard Medical School assigned a new function to APC in mature neurons, the ability to prevent axon growth. But the paper didn't explain the exact mechanisms. With a hunch that Id2 might play a role, Lasorella and Iavarone collaborated with the Harvard group. They showed that the mutated 'super Id2' protein allows axons to grow in different types of neurons, even in the presence of the myelin components that normally coat nerves and prevent them from regenerating after injury.

"This was a very exciting journey," says Iavarone. "We moved from finding Id2 as a target of the APC to showing there are APC–Id2 complexes in neurons, and we discovered how all of this happens." The path taken by these researchers, from looking at a protein complex associated with cell division to finding out how the same complex can affect axonal growth, shows how simple problems in molecular biology can turn into long, unexpected trips. ■

KEY CONTRIBUTOR

During the past ten years, Shuya Fukai, an associate professor at the Tokyo Institute of Technology, Japan, has developed techniques to help other scientists better determine three-dimensional protein structures, particularly for protein–RNA complexes. He began this work as a graduate student at the University of Tokyo, working with Osamu Nureki. This work, followed by continuing collaboration, produced clear snapshots of a

three-step chemical reaction (see page 419). According to Nureki, Fukai's expertise in crystallography, combined with his skills in biochemistry and molecular biology, made solving these structures possible.

Fukai plays down his role. "When someone meets a problem in crystallography, I give advice to help solve it, or, in more difficult cases, try to solve it myself," he says. "In this paper, as first author

Tomoyuki Numata is a good crystallographer, minimum advice was mostly sufficient."

Nureki demurs, saying that Fukai was instrumental not just in resolving the initial structures, but also in validating them using electron density maps, improving on the initial picture quality. These efforts paid off. "The three crystals correspond to the three reaction stages," Fukai says. "That's an unexpected success." ■